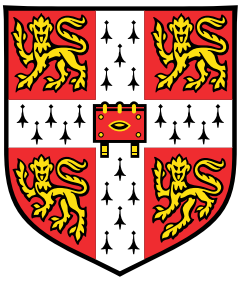
Statistical analysis of renal function in patients with cancer



Edward H. Williams

Cancer Research UK Cambridge Institute
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Clare Hall College

October 2019

Declaration

The contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. This dissertation contains fewer than 60,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Edward H. Williams
October 2019

Acknowledgements

I would first like to thank my PhD supervisors Simon Tavaré and Tobias Janovitz for their ongoing support and guidance throughout the last four years, and I'd like to extend my gratitude to their families for their hospitality when I visited them in the US. Thank you also to Andy Lynch for his advice during my time at Cambridge.

This work would not have been possible without those who helped collect patient data. In particular, I would like to mention Harry Potts who, whilst completing his medical studies, spent his free time collecting much of the data used for this thesis. Thank you also to Claire Connell and Cameron Whitley.

Thank you to those who have been sources of support, inspiration and guidance at the CRUK institute, especially Meltem Gürel, Sam Abujudeh, Henry Farmery, Ioana Olan and Ann Kaminski.

I would then like to thank my friends and family for their endless support, encouragement and faith in me. Thank you Ben, Harry, Kate, Rob, my Dad and in particular, my Mum.

Finally, thank you to my soon-to-be wife, Samantha Raven, whose endless support, understanding and patience made this PhD possible.

Abstract

This thesis presents statistical modelling approaches and analyses for the filtration function of the kidney in patients with cancer, using widely available clinical data.

The filtration function of the kidney is reported as the glomerular filtration rate (GFR). Many clinical decisions, for example dose calculations for chemotherapy drugs, are based on a patient's GFR. The most accurate way to determine a patient's GFR involves costly, time-consuming, and consequently frequently unavailable methods. Therefore, GFR is commonly estimated using statistical models based on routine clinical data. Although these models are used in patients with cancer, most of them have been developed using data from non-cancer patients with impaired kidney function.

Here we present a new statistical model for GFR in patients with cancer. We initially collected data from 2,579 patients managed at two cancer centres. Using regression modelling based on patient demographics and serum creatinine levels, a new model, CamGFR, was developed that is more accurate than other currently available models for GFR.

Following this, data from a further 7,944 patients were collected across nine cancer centres. These data enabled the assessment of the impact of different creatinine measurement methods, such as standardising creatinine to isotope dilution mass spectrometry (IDMS). Given that the method for creatinine measurement affects the reported laboratory value, the CamGFR model was adjusted to allow use for creatinine, which was or was not IDMS standardised.

Further analyses included the search for clinical correlates with renal function and assessment of whether such correlates can improve the estimation of GFR; the effect of treatment on GFR and estimated GFR; the examination of the effect of unstable creatinine on GFR estimation using longitudinal data; and a comparison of renal function between patients with and without cancer.

In summary, this thesis presents results that may help improve the estimation of kidney function and thereby care for patients with cancer.

Table of contents

List of figures

List of tables

Nomenclature

Acronyms / Abbreviations

ALP Alkaline phosphatase

ALT Alanine aminotransferase

AM Additive model

APTT Activated partial thromboplastin time

BSA Body surface area

CA125 Cancer antigen 125

CI Confidence interval

CKD-EPI Chronic Kidney Disease Epidemiology collaboration

CorCalcium Calcium corrected for albumin

CRP C-reactive protein

FAS Full age spectrum

GAM Generalised additive model

GFR Glomerular filtration rate

IDMS Isotope dilution mass spectrometry

IgA Immunoglobulin A antibodies

IgG Immunoglobulin G antibodies

IgM Immunoglobulin M antibodies

IQR Interquartile range

LD Lactate dehydrogenase

MCHC Mean corpuscular haemoglobin concentration

MCH Mean corpuscular haemoglobin

MCV Mean corpuscular volume

MDRD Modification of Diet in Renal Disease

MPV Mean platelet volume

PCT Procalcitonin

PDW Platelet Distribution Width

PLT Platelet count

PT Prothrombin time

RBC Red blood cell count

RDW Red cell distribution width

RMSE Root-mean-squared error

SUN Serum urea nitrogen

TSH Thyroid-stimulating hormone

WBC White blood cell count

Chapter 1

Introduction

1.1 Motivation

In clinical practice, decisions about patient care are made, in part, through information obtained from a patient's medical history but they are also made through medical investigations. These investigations often deliver biochemical results for individuals which, when aggregated, can be used to develop statistical models to better inform clinical practice and patient management.

One example is the use of statistical models to estimate organ function; one such function that has long been modelled is the kidney. Knowledge of kidney function is critical to decision-making in clinical practice as it influences decisions on treatment regimens, disease diagnosis and progression, and drug administration. The kidney has a further advantage over other organs in that it has a measurable quantity which represents its filtration function - glomerular filtration rate (GFR).

GFR is the rate at which fluid is filtered from the blood to the urine, which is the key role of the kidney. It has various uses in clinical decision-making and patient-management, one of the main ones being to diagnose chronic kidney disease (CKD). From a cancer perspective, a key use case of patient GFR is in dosing certain chemotherapy drugs.

Accurate assessment of a patient's GFR can be made using methods that measure the decay of exogenous filtration markers. However, these procedures are typically costly and require time and expertise. Hence, the overwhelming majority of reported GFR values are estimated from a

statistical model. These statistical models commonly relate GFR to the patient's blood serum creatinine concentration along with other demographic variables, such as age or sex.

There are existing statistical models to estimate GFR, however, there is cause to review them for the following key reasons. First, an increasing amount of data is being generated by hospitals during patient treatment, thus enabling improved statistical analysis and thus the potential for improved models. Secondly, the methods used to measure clinical variables, such as creatinine, are consistently evolving and improving, and this can affect any model based upon these variables. Finally, and specifically for patients with cancer, existing models to estimate GFR have been developed in non-cancer patient populations with limited large-scale validation for patients with cancer. The broad implications of this are currently unknown.

Considering the aforementioned points, the purpose of this thesis is to develop a model to best estimate GFR with a particular focus on patients with cancer. For this we collected a large multicentre dataset of patients with cancer and used this data not only to develop a new model but also to investigate additional aspects of patient biology that may inform renal function.

1.2 Overview of thesis

1.2.1 Chapter overview

In Chapter 2 we provide an introduction on the physiology of the kidney and approaches that have been taken to measure and estimate glomerular filtration rate (GFR). The chapter begins with an overview of renal function and its importance in patient management within clinical medicine. We then introduce the anatomy and physiology of the kidneys and the measure of a kidney's function through GFR. Next, this chapter defines GFR and its determinants, and provides methods used to both measure and estimate GFR. Chapter 2 concludes with a detailed discussion providing background on the development of both historical and current models used to estimate GFR, and critiques of these models. Whereby we establish a need for a multi-centre comparison of these models, and a new model, for GFR in patients with cancer.

In Chapter 3, we introduce the data for this thesis in addition to relevant clinical and statistical methodology. The first half of this chapter explains the clinical data: how data were acquired from cancer treatment hospitals or centres; the clinical methodology used to measure GFR

and the filtration marker, serum creatinine; the content of individual datasets; and finally how data were evaluated and quality controlled for the purpose of this thesis. In this initial section we point out the successive nature of data acquisition, with data acquired both pre- and post-publication of the initial CamGFR model, which is discussed further in Chapter 4. The second half of Chapter 3 then provides detail on statistical methodology that is relevant and will be used in this thesis, including regression modelling, model section methods and statistics used to compare different models.

The focus of Chapter 4 is on the development of a new serum creatinine-based model for GFR in patients with cancer, named CamGFR. This chapter is split into three component parts: the initial CamGFR development, comparison of creatinine values for different measurement methods and the resultant impact on modelling GFR, and finally, expansion of the CamGFR model for creatinine that is traceable to an IDMS standard.

In Chapter 5 we conduct an exploratory study of other biochemical and hematological blood tests aside from serum creatinine. In this study we use the additional blood tests to more accurately estimate GFR and to discuss the implications of the fitted relationships between GFR and particular blood tests.

In Chapter 6 we provide further statistical and exploratory studies, discussing additional clinical considerations for accurately modelling GFR in patients with cancer. We examine the effect of treatment, in particular chemotherapy, on a patient's GFR and creatinine, and the implications of this on estimating GFR. Using longitudinal serum creatinine data we then assess estimated GFR in patients whose serum creatinine is not stable and discuss future work in relation to this. Finally, we compare measured GFR in patients with cancer to healthy kidney donor patients.

In the concluding chapter, Chapter 7, we reflect upon the key findings in this thesis, consider limitations of this work and discuss how the CamGFR model and supporting studies may be developed in the future.

1.2.2 Original contributions

I have made a number of original contributions in this thesis. Particular contributions include:

- The first multi-centre comparison of models to estimate GFR in patients with cancer

- Development of a new model to estimate GFR in patients with cancer, for which the applicability may go further than only for patients with cancer
- An analysis of the relationship between measured GFR and commonly used biochemical and haematological blood tests

We expect to improve GFR modelling in the future by collecting further types of data. The initial stages of this work are outlined in Chapter 5 and Chapter 6.

1.2.3 Publications

During my time at Cambridge, I have been a named author of four publications. Two of these publications are directly related to the work for this thesis, which will be described in depth. These papers are listed below:

- Janowitz, T.[*], **Williams, E. H.**[*], Marshall, A., Ainsworth, N., Thomas, P. B., Sammut, S. J.,Shepherd, S., White, J., Mark, P. B., Lynch, A. G., Jodrell, D. I., Tavaré, S., and Earl, H. (2017). New model for estimating glomerular filtration rate in patients with cancer. Journal of Clinical Oncology, 35:2798–2805.

  [*] These authors contributed equally to the work.

- **Williams, E. H.**, Connell, C. M., Weaver, J. M. J., Beh, I., Potts, H., Whitley, C. T., Bird,N., Al-Sayed, T., Monaghan, P., Fehr, M., Cathomas, R., Bertelli, G., Quinton, A., Lewis, P., Shamash, J., Wilson, P., Dooley, M., Poole, S., Mark, P. B., Bookman, M. A., Earl, H., Jodrell, D., Tavaré, S., Lynch, A. G., and Janowitz, T. (2019). Multicentre validation of the CamGFR model for estimated glomerular filtration rate. JNCI Cancer Spectrum, 10.1093/jncics/pkz068.

Additionally, I have been involved in several other projects which are not discussed in this thesis. In particular, I was the statistician on an analysis of trends in T-cell checkpoint-targeted cancer immunotherapy clinical trials and on the analysis of interim results from the national Hepatitis Prevention, Control, and Elimination (HPCE) in Mongolia. This work contributed to the publication of two additional papers and a further paper which is currently under review. These papers are included in the Appendix and are listed below:

- Connell, C. M., Raby, S., Beh, I., Flint, T. R., **Williams, E. H.**, Fearon, D. T., Jodrell, D. I., Janowitz, T. (2017), Cancer immunotherapy trial registrations increase exponentially but chronic immunosuppressive glucocorticoid therapy may compromise outcomes. Annals of Oncology, 28:1678–1679.

- Connell, C. M., Raby, S., Beh, I., Flint, T. R., **Williams, E. H.**, Fearon, D. T., Jodrell, D. I., Janowitz, T. (2018). Cancer immunotherapy trials underutilize immune response monitoring. Oncologist 23:116–117.

- Dashdorj, N., Potts, H., Bungert, A., Bat-Ulzii, P., Dashtseren, B., Baatarsuren1, U., Myanganbayar, M., **Williams, E. H.**, Glenn, J. S., Genden, Z., Mordorj, A., Duger, D., Unurjarga, T., Yagaanbuyant, D., Janowitz, T., Dashdor N. (2019), Interim results of the Hepatitis C Elimination Program in Ulaanbaatar, Mongolia. Submitted to Gastroenterology

Chapter 2

Background to renal function and GFR

The aim of this chapter is to provide background information about glomerular filtration rate (GFR), including its physiology, definition, determinants, and the methods used to measure or estimate it. We begin in Section 2.1 by introducing renal function and establishing its importance for patient management within clinical medicine. In Section 2.2, we provide background information on the physiology of the kidney, a biological and mathematical definition of GFR, and normal ranges for GFR and discuss factors which affect these. In Section 2.3 we discuss measuring GFR using the clearance of an ideal filtration marker and its challenges. The chapter concludes with Section 2.4, where we discuss the theory behind estimating GFR using endogenous filtration markers, discuss the commonly used markers, and provide a background and critique for numerous previous models developed to estimate GFR.

In providing a background to GFR, this chapter uses several sources. Key sources are listed below and can be consulted for additional detail on the given subject areas:

- *Guyton and Hall: Textbook of Medical Physiology*, Chapter 26, 27 and 32 provides further detail on the physiology of the kidney [50]
- *Mathematical Physiology II: Systems Physiology*, Chapter 17, provides a more mathematical discussion of GFR [68]
- *Chronic Kidney Disease, Dialysis, and Transplantation*, Chapter 41 provides details on the methods and theory of measuring or estimating GFR [53]

2.1 Renal function and its role in patient management

The failure of one or more organ systems often leads to poor clinical outcomes for patients. One of the key organ systems that affects a patient's outcome is the renal system, which maintains many aspects of homeostasis in the body through regulation of electrolytes, water, and pH, excretion of metabolic waste products, and the production of hormones which control the production of red blood cells, regulate blood pressure, and regulate calcium levels in the body. Acute kidney injury or chronic kidney disease can cause dysregulation of electrolyte, water and pH levels, a build-up of metabolic waste products in the blood, and failure of the hormone systems regulated by the kidney. These problems are associated with poor clinical outcome for patients [129]

In healthy individuals and throughout most of our lives, renal function is maintained at a steady level, excluding changes due to circadian rhythm [69]. This steady level reflects healthy organ tissues and, upon injury, recovery of organ tissues. However, renal function can be affected by pathological disease processes, by ageing, and by iatrogenic insults for example surgery, radiocontrast agents or medications [1, 49]. Epidemiological studies have demonstrated that chronic kidney disease reduces life expectancy [41]. Loss of renal function is also associated with poor clinical outcomes in the intensive care unit (ICU) setting [125]. This is irrespective of the underlying disease process that initiated admission to the ICU. Importantly, a reversal of loss of renal function has been demonstrated as a successful way to prolong life, giving way to the common practice of renal replacement therapy, as haemofiltration in the ICU and as dialysis in outpatient management. A final piece of evidence for the importance of intact renal function is the survival benefit for patients who have received transplantations of donor kidneys [82].

Considering the fundamental importance of renal function to the wellbeing and survival of patients, it is unsurprising that clinicians use renal function as a guide for many important decision-making processes. Examples include, but are not restricted to: decision-making around nephrotoxic intravenous contrast media administration in radiological imaging techniques, such as computed tomography (CT) or magnetic resonance imaging (MRI) scans, as seen in the guidelines from the Royal College of Radiologists [124]; management through intensive care admission and haemofiltration; administration of dialysis or renal transplantation as mentioned above; and administration of potentially nephrotoxic drugs or of drugs that are filtered and preferentially excreted by the renal system and therefore have pharmacokinetics that are

dependent on renal function, such as carboplatin whose dose in calculated using the Calvert equation [13].

In conclusion, the assessment of the filtration function of the kidney is a fundamental component in the assessment of the physiological wellbeing of a patient and informs multiple critical decision-making processes for the management of patients and the prescription of many medications. Having established the principal importance of renal function, the next section will discuss the physiology of renal filtration, followed by a description of the methods by which renal function can be measured and finally, how renal function is commonly assessed in the clinical setting.

2.2 Anatomy and physiology of the kidney

The renal system comprises two kidneys located in the lower abdomen. They are well-vascularised retroperitoneal organs. Functionally, their role can be divided into metabolic contributions, endocrine functions and filtration functions. Metabolically, they have a role in the metabolism of carbohydrates, proteins and lipids, and in converting a precursor of vitamin D to its active form, calcitriol [26]. Endocrinologically, one of their most important functions is the production of erythropoietin, one of the hormones that regulates production and differentiation of blood precursor cells, in particular erythrocytes [61]. However, physiologically, the main function of kidneys is filtration, which maintains fluid/water and electrolyte homoeostasis in the entire organism and allows excretion of waste products [50, Chapter 26 and 32]. Henceforth, when referring to renal function, only the filtration function of the kidney is being considered.

The functional unit of renal filtration is the nephron, Figure 2.1 A. Each kidney is composed of millions of nephrons acting in parallel. These, in turn, consist of a glomerulus encased by a Bowman's capsule and a renal tubule. The glomerulus is a system of parallel branching and rejoining capillaries. Blood enters the glomerulus where, due to the hydrostatic and colloid osmotic forces acting across the capillary membrane [50, Chapter 26], fluid is filtered into the Bowman's capsule along with any dissolved substances of small molecular weight to form primary urine [68, Chapter 17]. A healthy individual produces up to 160 litres of primary urine per day. This is concentrated in the tubular system through processes of reabsorption coupled with electrolyte balancing to produce the secondary urine of about 1.5 litres that is collected in the bladder and ultimately excreted from the body.

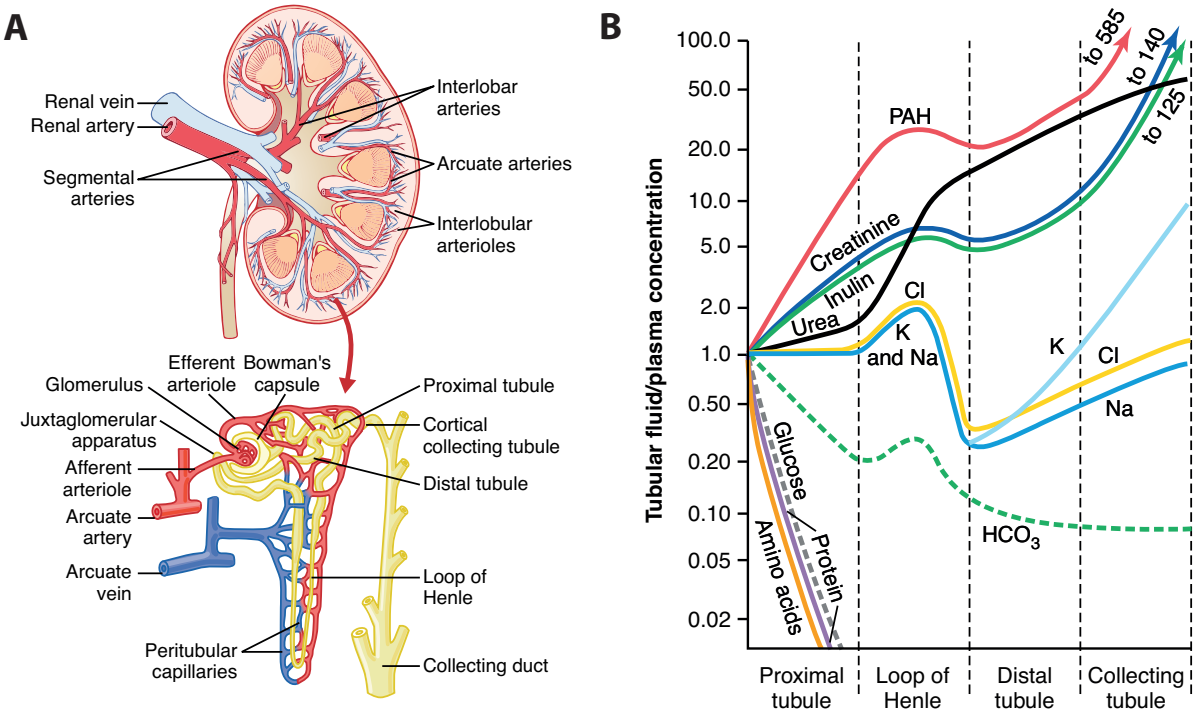


Figure 2.1 (A) Cross-section of the kidney and a schematic of a single nephron with the major components labelled. (B) Plot of average relative concentrations of various electrolytes and other metabolites as a function of distance along the renal tubule system. These include amino acids, protein, glucose, bicarbonate (HCO_3), sodium (Na), chlorine (Cl), potassium (K), urea, inulin, creatinine and polycyclic aromatic hydrocarbons (PAH). A value above 1 indicates that the substance is reabsorbed into the blood to a lesser extent than the water and/or is secreted into the urine in the tubules [50, Figure 26-3, p. 305 and Figure 27-14 p. 334].

Electrolytes and other metabolites are reabsorbed throughout the infrastructure of the nephron, with different metabolites reabsorbed at different stages, as shown in Figure 2.1 (B). This reabsorption is achieved by both passive and active transport. Amino acids, proteins and glucose, which are essential molecules for energy and tissue metabolism and maintenance of the body, are fully reabsorbed in the proximal tubule. Electrolytes, mainly potassium, sodium and chloride, are reabsorbed in the loop of Henle and then become more concentrated as water is reabsorbed in the distal tubule and collecting tubule, and are excreted in small amounts. The same applies to bicarbonate, which is the key substance by which the kidney regulates the pH (acid-base balance). Metabolic breakdown products such as creatinine and urea are increasingly concentrated throughout the process of water reabsorption and the formation of secondary urine [50, Chapter 27].

2.2.1 Definition of glomerular filtration rate

A measurable component of the filtration function of the kidney is the glomerular filtration rate (GFR). GFR is defined as the total volume of fluid filtered from the glomerular capillaries to the Bowman's capsules per unit time. This rate is influenced by hydrostatic and colloid osmotic pressure differences between the glomerular capillaries and the Bowman's capsule. Hence, a mathematical expression for the glomerular filtration rate would be:

$$GFR = K_f \times (P_G - P_B - \pi_G + \pi_B) \tag{2.1}$$

where K_f is the glomerular capillary filtration coefficient, a product of the filtering surface area and the hydraulic permeability of the capillaries, P_G is the glomerular hydrostatic pressure promoting filtration, P_B is the hydrostatic pressure in Bowman's capsule opposing filtration, π_G is the colloid osmotic pressure of the glomerular capillary plasma proteins opposing filtration and π_B the colloid osmotic pressure of the proteins in Bowman's capsule promoting filtration [50, Chapter 26]. However, this is a simplified relationship as the kidneys have mechanisms for autoregulation of GFR including tubuloglomerular feedback and the myogenic mechanism, which maintains a stable glomerular filtration rate despite changes in blood pressure.

2.2.2 Normal range and determinants of GFR

Before discussing normal ranges for GFR, we should first discuss the units typically used for GFR. Although GFR is measured as units of volume per unit time, e.g. millilitres per minute (mL/min), when assessing a patient's renal function, in particular with regards to chronic kidney disease, the effect of body size on GFR is often removed. This is achieved by normalising the GFR to a body surface area (BSA) of 1.73 m^2, where 1.73 m^2 is the estimated body surface area of an adult with a height of 1.7 m and a mass of 63 kg. As a result, GFR is commonly reported in the units mL/min/1.73m^2, rather than mL/min. It is important to note that BSA cannot be easily measured and is instead estimated using an equation involving patient height and weight e.g. DuBois & DuBois [33] or Haycock [52]. Some have questioned the rationale of normalising the GFR to body surface area due to inaccuracies of the estimation equation and the validity of representing body size using BSA [42]. This has led to suggestions for the use of extracellular fluid volume (ECFV) as opposed to BSA.

Numerous studies have been performed to calculate a normal range for GFR in healthy populations. In an analysis of 3,000 healthy potential kidney donors by The Renal Association they report an average GFR of 100 $\mathrm{mL/min/1.73m^2}$ for men and 98 $\mathrm{mL/min/1.73m^2}$ for women who were younger than 34, with average GFR decreasing by 0.6 $\mathrm{mL/min/1.73m^2/year}$ for men and 0.8 $\mathrm{mL/min/1.73m^2/year}$ for women after that [11]. Other studies have shown similar rates of decreasing GFR. However, there is not a consensus as to whether this decline is normal or due to age-related disease processes [53, Chapter 2].

Further to the above study, Delanaye et al. [30] discusses normal ranges from 23 different publications. These publications often stratified their results by patients' age and gender and were sometimes limited to a particular ethnic group or particular GFR measurement method. In these publications, the normal ranges differed considerably and differences between GFR adjusted by BSA in men and women were reported in some studies but not in others. When a difference was observed, this was typically around 8% in men and women of a similar age.

As well as body size and sex there are additional determinants of GFR, such as other patient demographics, protein intake, pregnancy, diurnal variation and antihypertensive therapy. The association between GFR and age has been shown in both cross-sectional and longitudinal studies. As stated above, The Renal Association reported a decline of 8 $\mathrm{mL/min/1.73m^2}$ and 6 $\mathrm{mL/min/1.73m^2}$ per decade after the age of 34 for men and women respectively [11]. Evaluating differences in ethnicity are complicated by the fact that most studies of normal GFR values used Caucasian populations. Where there has been evaluation of normal GFR in different ethnic groups, typically the ethnic populations were homogeneous and hence it is difficult to evaluate whether any observed differences are due to ethnicity or other study designs. One study of an Indian population gave an average GFR of 81.4 $\mathrm{mL/min/1.73m^2}$ for potential kidney donors [4] and in a study of a South Asian population, the average GFR was reported as 94.1 $\mathrm{mL/min/1.73m^2}$ [59].

The effects of protein intake on the GFR have been widely studied, especially in the context of chronic kidney disease. After consuming a protein meal GFR has been shown to rise, with the type of protein and feeding habits thought to cause varying effects [53]. GFR has also been observed to increase by up to 50% during pregnancy [18]. This increase is thought to be caused by a decrease in oncotic pressure, π_G, which itself is caused by an expansion of the plasma volume.

Diurnal or circadian variations in GFR are well established, with a difference of up to 30% over the 24 hour mean [138]. The drivers for this fluctuation are not fully understood and may be partly or wholly driven by diurnal variations in other factors affecting GFR such as protein intake, mentioned above, or hydration [3]. Daily fluctuations in blood pressure in normal conditions do not affect the GFR. This is due to renal autoregulation whereby the kidneys counteract the changes in blood pressure by constriction or dilation of the afferent arterioles, maintaining P_G. However, chronic elevation of blood pressure has been associated with deterioration in GFR, a phenomenon that is thought to be due to remodelling of the microvasculature of the kidney. This is true for seemingly isolated hypertension and has been confirmed for those conditions where hypertension co-exists or as a secondary phenomenon for a primary disease such as diabetes. However, diabetes in its own right causes injury to the vasculature presumably due to non-specific protein glycosylation by the hemiacetal form of glucose, and this in turn will cause hypertension as well as decreasing renal function.

2.3 Measuring GFR

Given the importance of GFR in clinical decision-making, considerable efforts have been dedicated to the assessment of GFR. GFR cannot be measured directly, so its value is instead inferred from the clearance of an ideal filtration marker, which ideally has the following four properties:

1. It is cleared from the blood by urinary excretion alone.
2. It is freely filtered from the glomerular capillaries into the Bowman's capsule.
3. It is not secreted, reabsorbed, synthesised or metabolised as filtrate flows through the renal tubules.
4. It does not affect the filtration function of the kidney.

If a substance meets all these requirements, then GFR is equal to the clearance rate C_x of the filtration marker x. Further, the amount which is eliminated from the plasma $(A_x = GFR \times P_x)$, will equal the amount of the substance which is excreted in the urine $(U_x \times V)$. Hence, we can

write GFR as:

$$GFR = C_x = \frac{U_x \times V}{P_x} \tag{2.2}$$

$$= \frac{A_x}{P_x} \tag{2.3}$$

where C_x is the clearance of filtration marker x, A_x is the amount eliminated from the plasma per unit time, P_x is the average plasma concentration, U_x is the average urine concentration, and V is the urine flow rate.

GFR can be measured in two ways using an ideal filtration marker. Traditionally, it would be measured via the urinary clearance of the filtration marker thus computing GFR using equation 2.2. Alternatively, GFR can be calculated using the plasma clearance of an exogenous filtration marker following a bolus intravenous injection, using equation 2.3 with P_x calculated using the area under the decay curve and A_x the amount administered. Plasma clearance has a key advantage over urinary clearance in that it does not require urinary collection. However, it still requires a relatively long waiting period to determine the decay curve, approximately 5 hours, which is exacerbated for patients with particularly low GFR values.

Many filtration markers have been proposed and used to measure GFR. Historically, GFR was measured with urinary clearance of endogenous filtration markers such as creatinine or urea. However, these markers cannot be truly considered ideal as they do not meet all the properties stipulated above. In particular, creatinine clearance systematically underestimates GFR due to creatinine reabsorption in the renal tubules. One of the early exogenous markers proposed was inulin, which meets all the requirements for an ideal filtration marker [53, Chapter 2]. However, it has the disadvantages of being difficult to dissolve in water, difficult to measure, in short supply, and it is required to be administered as a continuous infusion thus only allowing it to be used for urinary clearance calculations. Hence, other markers have been proposed and are now more commonly used, such as: ^{125}I-Iothalmate, Iohexol, ^{51}Cr-Ethylenediaminetetraacetic acid (^{51}Cr-EDTA) or ^{99m}Tc-Diethylenetriamine pentaacetate (^{99m}Tc-DPTA). A comparison of these markers can be found in [53, Chapter 2].

2.4 Estimating GFR

The clearance methods for measuring GFR described above are costly, challenging to perform and time-consuming, particularly for regular clinical practice. As an alternative, regression-based models have been developed which regress GFR against the plasma or serum level of a naturally occurring filtration marker along with other variables known to affect its level, such as the factors discussed in Section 2.2.2. Most commonly, this filtration marker is chosen to be serum creatinine. However, urea, cystatin C or a combination of markers have also been used.

2.4.1 Theoretical rationale

Equation 2.2 for GFR was derived from equating the plasma clearance rate of a filtration marker to the rate at which it is excreted in the urine. If the filtration marker is also at a steady state in the plasma, then its generation rate G_x must also be equal to:

$$G_x = GFR \times P_x \tag{2.4}$$

However, if the marker x is not a perfect filtration marker and some reabsorption or secretion occurs in the tubules (which is the case for the naturally occurring filtration markers mentioned above) this can be added to Equation 2.4 as follows:

$$G_x = GFR \times P_x - R_x + S_x \tag{2.5}$$

where R_x is the reabsorption rate and S_x is the secretion rate. Hence, once we rearrange this equation:

$$GFR = \frac{G_x + R_x - S_x}{P_x} \tag{2.6}$$

Assuming the generation, reabsorption and secretion rates remain the same, we observe that a change in the plasma concentration implies GFR has also changed. However, we cannot practically measure the generation, reabsorption and secretion rates in the above equation. Instead, we use patient demographic variables as surrogates for these rates and use regression analysis to relate GFR to the plasma concentration of the filtration marker and the demographic variables.

2.4.2 Filtration markers

The most commonly used filtration marker for regression-based models for GFR is serum creatinine. Creatinine is formed by a non-enzymatic cyclic reaction of creatine or phosphocreatine and as such can be understood as a by-product of energy maintenance and metabolism, which occurs mostly in the muscles [53, Chapter 2]. This means that creatinine is produced at a reasonably constant rate, which depends mainly on muscle mass but to a lesser extent on diet, creatine supplements and disorders of skeletal muscle. As creatinine has a small molecular diameter and does not bind to plasma proteins, it is freely filtered from the glomerular capillaries into the Bowman's capsule. Creatinine is also secreted by the tubules (see Figure 2.1), thus creatinine clearance is an overestimate of GFR.

Creatinine can be measured easily in serum, plasma and urine by a variety of different methods. The most accurate method to measure creatinine uses isotope dilution mass spectrometry (IDMS), using either gas or liquid chromatography [131]. However, this method is not practical on a large scale. Instead, one of two groups of methods are usually used, Jaffe or enzymatic. Jaffe (or alkaline picrate) methods are the more traditional methods used to measure creatinine. They use the Jaffe reaction where creatinine reacts with picrate ions under alkaline conditions to form a red-orange complex which can be detected and quantified [57]. However, this reaction is not entirely specific, as picrate can react with other components, known as pseudo-chromogens. The assay is also known to have interaction with bilirubin or other specific drugs [28]. This interference causes conventional Jaffe assays to overestimate the concentration of serum creatinine in a given sample. The kinetic Jaffe method is an adjustment to the above where the colour change is measured kinetically. This reduces the interference from other molecules due to the differential rate of colour development for non-creatinine substances compared to creatinine. Enzymatic methods were developed to further reduce the interference from other substances. The two enzymatic methods used in clinical laboratories are the creatinine iminohydrolase method, and the creatininase, creatinase, and sarcosine oxidase method [53, Chapter 2].

Both Jaffe and enzymatic methods are known to have inter-laboratory variation, which is problematic when using creatinine as a marker for kidney function [23]. Hence the National Kidney Disease Education Program (NKDEP) and the International Federation of Clinical Chemistry and Laboratory Medicine produced guidelines that recommended standardisation of all creatinine assays to IDMS-traceable reference values [84]. IDMS-traceable creatinine is the standard in the US and is now becoming common in the UK.

The other commonly used filtration marker is cystatin C, with reports that it should be used as an alternative, or in addition to, creatinine. Cystatin C is a small, 122 amino acid, 13 kDa protein that is a member of the family of cysteine proteinase inhibitors [72]. Its primary biological function is to regulate the activity of cysteine protease enzymes, which degrade proteins. Cystatin C is thought to be produced by all human nucleated cells at a stable rate and as such is found in virtually all cells and body fluids. Due to its small size (3nm) it can be freely filtered by the glomerulus. Once filtered, it is not secreted but is reabsorbed in the tubules and degraded by the proximal tubular cells, thus not returning to the blood. A study which used labelled cystatin C in rats showed that the clearance rate was 94% that of ^{51}Cr-EDTA and it was highly correlated with ^{51}Cr-EDTA as renal function was artificially reduced [122]. The study also observed extrarenal elimination of cystatin C in the spleen, diaphragm, heart, liver, and lungs at approximately 15% of total cystatin C elimination.

There are two methods by which commercially available assays measure cystatin C: particle-enhanced turbidimetric immunoassay (PETIA) or particle-enhanced nephelometric immunoassay (PENIA). Considerable differences have been observed between these two methods in several studies [128, 29], highlighting the need for assay standardisation. In light of this, an International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) Working Group on Standardization of Cystatin C (WG-SCC) was established, and human serum reference material for cystatin C has been developed [47].

The use of cystatin C as a filtration marker over creatinine has several key benefits. The main benefit of cystatin C is that it is not as dependent on body composition, in particular muscle mass has little or no effect on its concentration [127, 115]. This means that models using cystatin C have less of a need to factor in the patient's sex or race when modelling GFR. Furthermore, this lack of dependence on body composition allows the possibility of producing a model for both adult and paediatric patients. Cystatin C based models have been shown to better predict the consequences of kidney disease, in particular end-stage renal disease, cardiovascular manifestation, hospitalisation and death [46].

However, there are also disadvantages of cystatin C. Studies have found that cystatin C may be affected by a number of pathological conditions, which may be caused by an underlying change in the patient's renal function or may be independent of the renal function. Specific examples include increased cystatin C for patients with HIV infection [87], asthma [20], hyperthyroidism [39], and corticosteroid treatment [8]. Decreased cystatin C was observed in abdominal aortic aneurysm [119], neurologic inflammatory disorder [85], and (in particular for this thesis) some

cancers [70, 130]. Furthermore, another disadvantage of cystatin C is the high cost of measuring it compared with creatinine. In 2013 the reagent cost per test was $4 for cystatin C, compared with $1.50 for enzymatic creatinine and $0.20 for Jaffe creatinine [120]. This high cost also means that cystatin C is not currently routinely measured at hospitals; this is in particular true for the hospitals where data were acquired for this thesis. Thus for a hospital to change from a creatinine based model to a cystatin C based model they would need to add cystatin C to the routine biochemical measurements performed, at further expense.

Another filtration marker which has been used to assess renal function is serum urea nitrogen (SUN). SUN concentration in urine was used as an assessment of renal function before regression-based models for GFR were developed. SUN is a breakdown product in the nitrogen cycle, which is made in the liver and is mainly excreted through the kidneys. Its concentration in plasma is affected by numerous factors, in addition to GFR. These factors include dietary protein intake, corticosteroids, diuretics, or tetracyclines, infection, trauma, congestive heart failure, and sodium depletion [53, Chapter 2]. However, even given these factors, the concentration of SUN is predictive of a patient's renal function, with its concentration correlated with serum creatinine concentration. Compared with creatinine, less effort has been made to study and resolve any inter-laboratory variability of the measurement of SUN.

2.4.3 Previous models

Numerous models have been developed that model GFR and are used to estimate GFR in clinical practice. This section will provide details of the more commonly used models both past and present. We focus on creatinine-based models which are used later in this thesis. However, some cystatin C models are also briefly described. The variables and units used for each model are consistent with units provided is Table 2.1. Note that to maintain the consistency for units, some models may be displayed differently from their original publication.

Cockcroft-Gault

In 1976, Cockcroft and Gault published a model to estimate GFR which is still used to the present day. The Cockcroft-Gault model used serum creatinine, along with age, sex, weight and height [22]. This equation was developed to estimate creatinine clearance, which is an overestimation of GFR, as mentioned in Section 2.4.2. The model was derived using data from

Key	Unit	Variable description
GFR	mL/min	Glomerular filtration rate
GFR_{BSA}	$mL/min/1.73m^2$	Body surface area adjusted glomerular filtration rate
CCr	mL/min	Creatinine clearance
CCr_{BSA}	$mL/min/1.73m^2$	Body surface area adjusted creatinine clearance
SCr	mg/dL	Serum creatinine
SUN	mg/dL	Serum urea nitrogen
Alb	g/dL	Serum albumin
CysC	mg/L	Cystatin C
Age	years	Age
Ht	cm	Height
Wt	kg	Weight
BSA	m^2	Body surface area
Sex_F	Male = 0, Female = 1	Categorical variable specifying gender
Black	Ethnically black = 1 Other = 0	Categorical variable specifying whether the patient is ethnically black
NSAID	Yes = 1, No = 0	Non-steroidal anti-inflammatory drugs

Table 2.1 Model variable notation and units.

249 patients and has the following form:

$$\mathrm{CCr} = \frac{0.01131\mathrm{Wt}\,(140-\mathrm{Age})\,(1-0.15\mathrm{Sex_F})}{\mathrm{SCr}}$$

Compared with previously published models [34, 62, 63] the Cockcroft-Gault model showed improved accuracy in estimating creatinine clearance and was accepted as the standard method to estimate creatinine clearance for many years.

The Cockcroft-Gault equation is still used today in certain clinical contexts due to customary clinical practice. This is despite the fact that there are now more accurate models available and many inherent problems associated with the Cockcroft-Gault equation. Cockcroft-Gault was derived using a dataset of 249 patients who were mostly hospitalised men, all of whom had chronic kidney disease (CKD). This fact alone would indicate that there would be limitations in using this equation when the patient does not have CKD because they are more likely to have a higher GFR. Additionally, it was developed using older methods to measure serum creatinine, which have since improved in both accuracy and reliability.

MDRD

The Modification of Diet in Renal Disease (MDRD) study was a multicentre trial to evaluate the effect of dietary protein restriction and strict blood pressure control on the progression of renal disease [74]. The study comprised 1,628 CKD patients, with measurements for GFR, serum creatinine, and several other variables. A secondary purpose of this trial was to develop an equation to improve the estimation of GFR from serum creatinine.

In the study, the accurate GFR measurement was calculated as the renal clearance of ^{125}-I Iothalamate. GFR was expressed per $1.73m^2$ and modelled on the log scale. The study used a stepwise linear regression to determine a set of variables that jointly predict GFR. The variables considered were: weight, height, sex, ethnicity, age, diagnosis of diabetes, serum creatinine concentration, serum urea nitrogen level, serum albumin level, serum phosphorus level, serum calcium level, mean arterial pressure, urine creatinine level, urine urea nitrogen level, urine protein level, and urine phosphorus level. The model was developed using a random training set of 1,070 out of the 1,628 patients.

From the MDRD study, a six-variable equation was developed containing the variables serum creatinine, age, serum urea nitrogen, serum albumin, sex and ethnicity (black vs. non-black):

$$\begin{aligned} \mathrm{GFR_{BSA}} = 170\,\mathrm{SCr}^{-0.999}\mathrm{Age}^{-0.176}\mathrm{SUN}^{0.170}\mathrm{Alb}^{0.318} \\ \times (1-0.248\mathrm{Sex_F})(1+0.180\mathrm{Black}) \end{aligned} \tag{2.7}$$

However, due to serum albumin level and serum urea nitrogen level often not being available in clinical practice, a simpler equation which did not contain these two variables was produced:

$$\mathrm{GFR_{BSA}} = 186\,\mathrm{Age}^{-0.203}\mathrm{SCr}^{-1.154}(1-0.268\mathrm{Sex_F})(1+0.212\mathrm{Black}) \tag{2.8}$$

After the National Kidney Disease Education Programme (NKDEP) published recommendations for the measurement of creatinine in 2006 [84], which resulted in isotope dilution mass spectrometry traceable (IDMS-traceable) creatinine being adopted, the MDRD equation was subsequently updated for this creatinine type [76]. In this adjustment, the constant of proportionality was changed from 186 to 175, resulting in the following equation:

$$\mathrm{GFR_{BSA}} = 175\,\mathrm{Age}^{-0.203}\mathrm{SCr}^{-1.154}(1-0.268\mathrm{Sex_F})(1+0.212\mathrm{Black}) \tag{2.9}$$

This change improved the performance of the MDRD equation when creatinine was IDMS-traceable, resulting in less biased estimates [19]. As with the Cockcroft-Gault equation, the MDRD equation also has the limitation of being developed using a cohort of only CKD patients. So, once again, this equation is likely to be inaccurate for any patient who does not suffer from CKD.

CKD-EPI

Currently, the most commonly used creatinine-based model for GFR is the CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration) published in 2009 [77]. It was developed with the aim of creating a more accurate equation than the MDRD and to remove the systematic underestimation of GFR at high values. A pooled multicentre dataset of 8,254 patients from six research studies and four clinical populations was used in the development of the equation. It was further validated with an external dataset of 3,896 from six research studies and ten clinical populations. These patients were a mix of of high-risk populations (defined as GFR $<90\,\mathrm{mL/min/1.73m^2}$), such as patients with CKD, and low-risk populations who were potential kidney donors. Of the 8,254 in the development and internal validation dataset, 71% came from high-risk populations. Of the 3,896 validation samples, 72% came from high-risk populations. For the development and internal validation datasets, the accurate measurement of GFR was calculated using the urinary clearance of ^{125}I-Iothalamate. However, for the external validation dataset, iothalamate and other exogenous markers were used. The original serum creatinine values were recalibrated to the Roche enzymatic method [77].

The equation developed has the form:

$$\begin{aligned} \mathrm{GFR_{BSA}} = 141 \min\left(\frac{\mathrm{SCr}}{\kappa}, 1\right)^{\alpha} \max\left(\frac{\mathrm{SCr}}{\kappa}, 1\right)^{1.209} \\ \times \mathrm{Age}^{0.993}(1 + 0.159\mathrm{Black})(1 + 0.018\mathrm{Sex_F}) \end{aligned} \tag{2.10}$$

where

κ is 0.7 for females and 0.9 for males
α is -0.329 for females and -0.411 for males

This equation can be simplified to a linear regression for $\log(\mathrm{GFR_{BSA}})$ with the dependent variables log(SCr) and log(Age), sex and race. The fitted relationship for log(SCr) is continuous

piecewise linear with one spline point, which is at a different value of log(SCr) for males and females (Figure 2.2). The CKD-EPI model was compared to the MDRD model and showed improved accuracy in both the internal and external validation datasets, with results consistent across different subgroups of the explanatory variables [77].

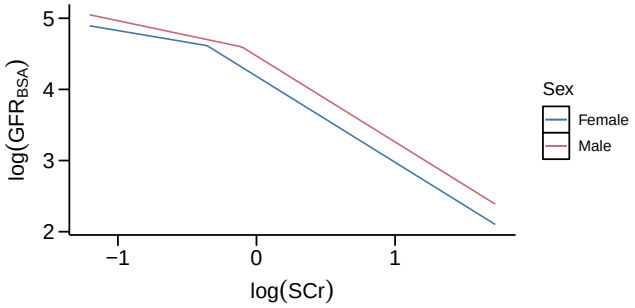


Figure 2.2 The fitted relationship between $\log(\text{GFR}_{\text{BSA}})$ and $\log(\text{SCr})$ in the CKD-EPI model. Specific examples shown are for a 50 years old male and female with no race adjustment.

Other models

Below is a list of other models; this is not an exhaustive list, but includes all models which were analysed in this thesis, along with more recent cystatin C based models. In addition to the models listed here, others have been developed or are a reparametrisation for different subgroups of patients [78, 81, 54, 123]. A comprehensive list of 26 creatinine based models published prior to 2012 was composed by Shaffi et al. [116], which compared these models for solid organ transplant patients.

1. **Jellife** [62, 64, 63] Jellife published several models between 1971 and 1973 concerning models to estimate GFR based on serum creatinine measurements. The most used model was the 1972 publication, which was developed using 250 observations from 8 men and 6 women [63]:
$$\text{CCr}_{\text{BSA}} = \frac{98 - 0.8(\text{Age} - 20)(1 - 0.9\text{Sex}_{\text{F}})}{\text{SCr}} \tag{2.11}$$
Similar to the Cockcroft-Gault model mentioned previously, this model has many of the same issues and so its use has declined over time in favour of newer models.

2. **Wright** [137]: Wright et al. published a set of models in 2001 which either did, or did not, include creatine kinase concentration as a variable and used either Jaffe serum

creatinine measurements or enzymatic serum creatinine measurements. This is one of only two models developed using data specifically from patients with cancer and the only model out of the two to be widely used. The model was developed using a dataset containing only 62 patients. Below is the model that does not contain creatine kinase and uses Jaffe serum creatinine:

$$\text{GFR} = \frac{(6580 - 38.8\text{Age})\,(1 - 0.168\text{Sex}_\text{F})\,\text{BSA}}{\text{SCr}} \tag{2.12}$$

3. **Mayo quadratic** [108]: The Mayo quadratic model was developed after the publication of the MDRD model in an attempt to improve the estimation of GFR in patients with age-normal GFR. It is the only model to model the relationship between $\log(\text{GFR}_\text{BSA})$ as a quadratic function for the inverse of creatinine.

$$\text{GFR}_\text{BSA} = \exp\left(1.911 + \frac{5.249}{\text{SCr}} - \frac{2.114}{\text{SCr}^2} - 0.00686\text{Age} - 0.205\text{Sex}_\text{F}\right) \tag{2.13}$$

Additionally, for the above model, if $\text{SCr} < 0.8$mg/dL, use 0.8 for SCr.

4. **Lund-Malmo** [7]: This is a revised version of the original Lund-Malmo model [6]. It was developed using a Swedish population of 850 patients and used IDMS-traceable creatinine. Similar to the CKD-EPI model, the fitted relationship between $\log(\text{GFR}_\text{BSA})$ and creatinine is piecewise continuous curve. The model has the following form:

$$\text{GFR}_\text{BSA} = \exp(X + 2.5 + 0.06\text{Sex}_\text{F} - 0.0158\text{Age} + 0.438\log(\text{Age})) \tag{2.14}$$

where

$$\text{if female and SCr} < 150\text{: } X = 0.0121(150 - \text{SCr})$$
$$\text{if female and SCr} \geq 150\text{: } X = -0.926\log\left(\frac{\text{SCr}}{150}\right)$$
$$\text{if male and SCr} < 180\text{: } X = 0.00968(180 - \text{SCr})$$
$$\text{if male and SCr} \geq 180\text{: } X = -0.926\log\left(\frac{\text{SCr}}{180}\right)$$

5. **Giglio** [43]: This model was developed using 641 patients with cancer who were treated at Sahlgrenska University Hospital, Gothenburg, Sweden. Along with the Wright model discussed above, this was the only other model developed using patients with cancer and the only one to use a reasonably large population of patients with cancer. This model is

rarely discussed in the literature.

$$\begin{aligned}\log_{10}(\mathrm{GFR_{BSA}}) = 2.36 - 0.33\log_{10}(\mathrm{Age}) - 0.78\log_{10}(88.4\mathrm{SCr}) + \\ 0.87\log_{10}(\mathrm{Wt}) - 0.03\,\mathrm{NSAID}\log_{10}(88.4\mathrm{SCr}) + \\ \mathrm{Sex_F}(0.05\,\mathrm{NSAID} - 0.003\,\mathrm{BMI})\end{aligned} \tag{2.15}$$

6. **FAS** [102, 101]: The FAS equation was developed to be suitable for estimating GFR in all age groups. The model takes a slightly different approach by first normalising creatinine by the median creatinine for matched healthy subjects. The first model ($\mathrm{FAS_{crea}}$) is produced using creatinine and is shown below:

$$\mathrm{GFR_{BSA}} = \begin{cases} \frac{107.3}{\mathrm{SCr}/Q_{\mathrm{crea}}} & \text{for } 2 \leq \mathrm{Age} \leq 40 \\ \frac{107.3}{\mathrm{SCr}/Q_{\mathrm{crea}}} \times 0.988^{\mathrm{Age}-40} & \text{for } \mathrm{Age} > 40 \end{cases} \tag{2.16}$$

where Q_{crea} is the median serum creatinine in a healthy population with the same age and sex for the $\mathrm{FAS_{crea}}$ model. A table of Q_{crea} values was provided in the manuscript [102]. This group added to the $\mathrm{FAS_{crea}}$ model by producing two new models which either replaced creatinine with cystatin C ($\mathrm{FAS_{cysC}}$):

$$\mathrm{GFR_{BSA}} = \begin{cases} \frac{107.3}{\mathrm{CysC}/Q_{\mathrm{cysC}}} & \text{for } 2 \leq \mathrm{Age} \leq 40 \\ \frac{107.3}{\mathrm{CysC}/Q_{\mathrm{cysC}}} \times 0.988^{\mathrm{Age}-40} & \text{for } \mathrm{Age} > 40 \end{cases} \tag{2.17}$$

or used a weighted sum of the creatinine and cystatin C terms ($\mathrm{FAS_{combi}}$):

$$\mathrm{GFR_{BSA}} = \begin{cases} \frac{107.3}{\alpha\mathrm{SCr}/Q_{\mathrm{crea}} + (1-\alpha)\mathrm{CysC}/Q_{\mathrm{cysC}}} & \text{for } 2 \leq \mathrm{Age} \leq 40 \\ \frac{107.3}{\alpha\mathrm{SCr}/Q_{\mathrm{crea}} + (1-\alpha)\mathrm{CysC}/Q_{\mathrm{cysC}}} \times 0.988^{\mathrm{Age}-40} & \text{for } \mathrm{Age} > 40 \end{cases} \tag{2.18}$$

where typically $\alpha = 0.5$ is chosen.

7. **CKD-EPI - cystatin C** [58]: An extension of the CKD-EPI model which used cystatin C with or instead of creatinine and the filtration marker. The version below includes cystatin C only:

$$\begin{aligned}\mathrm{GFR_{BSA}} = 133\min\left(\frac{\mathrm{CysC}}{0.8}, 1\right)^{-0.498}\max\left(\frac{\mathrm{CysC}}{0.8}, 1\right)^{-1.328} \\ \times\,\mathrm{Age}^{0.996}(1 - 0.068\mathrm{Sex_F})\end{aligned} \tag{2.19}$$

The version which includes both cystatin C and creatinine takes a similar form and can be found in the publication [58].

8. **CAPA** [48]: The CAPA model was developed using the same Swedish patient population as the Lund-Malmo mentioned previously with the addition of some adults in Japan and children in Holland and Sweden. It uses cystatin C only as its filtration marker and is valid for both adults and children.

$$\text{GFR}_{\text{BSA}} = 130\text{CysC}^{-1.069}\text{Age}^{-0.117} - 7 \tag{2.20}$$

9. **Schwartz** [112] Originally developed in the 1970 [111, 110], the Schwartz model has since been revised using IDMS-traceable creatinine [112]. The model is suitable for children 1-18 years old. The model, also known as Bedside Schwartz, commonly takes the following form:

$$\text{GFR}_{\text{BSA}} = \frac{0.413\text{Ht}}{\text{SCr}} \tag{2.21}$$

A more accurate model which contains the additional variables serum urea nitrogen, cystatin C and height was also produced:

$$\begin{aligned}\text{GFR}_{\text{BSA}} = 0.391\,(1.099 - 0.099\text{Sex}_{\text{F}})\left(\frac{\text{Ht}}{\text{SCr}}\right)^{0.516} \\ \left(\frac{1.8}{\text{CysC}}\right)^{0.294}\left(\frac{30}{\text{SUN}}\right)^{0.169}\left(\frac{\text{Ht}}{140}\right)^{0.188}\end{aligned} \tag{2.22}$$

2.4.4 Comparison of models for patients with cancer

Some previous efforts have been made to compare these models in patients with cancer, particularly looking at patients who are treated with platinum-based chemotherapy drugs such as cisplatin or carboplatin. We searched for all such comparisons conducted since 2007 where different models to estimate GFR in an adult population were compared. We only included studies that compared estimated GFR to GFR measured using a nuclear tracer method. In total we reviewed eight publications, with the results from each publication summarised below.

- Lindberg et al. [80] compared Cockcroft-Gault and CKD-EPI for 94 patients with head and neck cancer who were treated with cisplatin. The publication found a bias for GFR estimated using CKD-EPI but not for Cockroft-Gault

- Cathomas et al. [15] compared Cockcroft-Gault, CKD-EPI, Jellife, Martin, Mayo, MDRD, Wright for 426 patients with stage I seminoma treated with adjuvant carboplatin treated between 1999 and 2012. Wright and Cockcroft-Gault showed the least bias of the seven models. All models were shown to lead to underdosing of carboplatin compared with ^{51}Cr-EDTA GFR. The data used in this publication were acquired for analysis in this thesis.

- Lauritsen et al. [73] compared Cockcroft-Gault, MDRD, Wright and CKD-EPI for 390 patients with disseminated germ cell cancer who received conventional chemotherapy between 1984 and 2007 at a single centre. For GFR prior to chemotherapy treatment, Cockcroft-Gault was the most accurate with a mean absolute percentage error (MAPE) of 10.8% followed by Wright at 11.1%. Wright was the least biased (based on percentage error). After chemotherapy, measured GFR was seen to decline significantly without a corresponding increase in creatinine. Thus all models overestimated GFR after receiving chemotherapy. This publication is discussed further in Section 6.1

- Quinton et al. [106] compared Cockcroft–Gault, Jelliffe and Wright for 68 patients with stage I testicular seminoma. The Wright model was found to be most accurate with a MAPE of 12.9%, and also the least biased. The data used in this publication were acquired for analysis in this thesis.

- Craig et al. [24] compared MDRD, CKD-EPI and Cockcroft-Gault for 288 patients who were treated with chemotherapy. All three models were found to be biased to overestimate GFR. Cockroft-Gault was the most accurate based on the proportion of patients with an absolute percentage error less than 30% (P30).

- Ainsworth et al. [2] compared Cockcroft–Gault, Jelliffe, Wright and MDRD in 660 oncology patients from Addenbrookes Hospital, Cambridge and found that Cockcroft–Gault was the most accurate and least biased. The patient data used in this publication form a subset of the data analysed acquired from Addenbrookes Hospital, Cambridge for this thesis.

- Barraclough et al. [5] compared Cockcroft–Gault, MDRD and Wright for 367 patients with cancer between April 2005 and January 2007. Wright was found to be the most accurate with a MAPE of 19%. Wright and Cockroft Gault were the least biased.

- Poole et al. [100] compared MDRD, Cockcroft-Gault, Wright, Martin, and Jelliffe in 510 patients with cancer. Wright was the most accurate as measured by the proportion

of patients with an estimated GFR within 30% of their measured GFR (P30). Wright and Martin were the only two models which were not biased. Data from this publication were acquired for analysis in this thesis.

This review demonstrated a need to carry out a large multi-centre validation of models to estimate GFR in patients with cancer for a number of reasons. First, although the Wright model was seen to be most accurate in three out of the eight publications, they generally had inconsistent conclusions in assessing which model is more accurate or least biased. Moreover, all of these publications used patients from a single centre, with some comprising only of patients with a particular cancer type. Finally, most publications used creatinine which was not IDMS-traceable.

A further demonstration of the inconsistent choice of model for patients with cancer can be shown by examining the methods used to determine GFR in clinical trials involving carboplatin. In a sample of clinical trials, ^{51}Cr-EDTA [56, 92, 140], the Cockcroft-Gault model [66, 27, 105], the Jelliffe model [32, 67, 95, 103], and 24 hour urine creatinine collections [92] were all used to calculate GFR. All these calculated the dose for carboplatin for the determined GFR using the Calvert equation [13].

2.5 Summary

This chapter has introduced GFR, its determinants and methods used to both measure and estimate it. We have discussed previous regression-based models to estimate GFR and identified a clinical need to develop a new model especially for patients with cancer. This need arises due to the fact that none of the available models, with the exception of the Wright model which was developed using a small sample of 62 patients and the Giglio model which is rarely cited, have been developed using data from patients with cancer. Despite this, these previous models are commonly used in day-to-day clinical medicine for patients with cancer, in particular for dosing carboplatin.

Additionally, there is limited consensus for the best method to estimate GFR when different models are compared in a population of patients with cancer. This is demonstrated in the literature review of previous publications. It is also further evidenced by the lack of consensus in the method used to calculate GFR with respect to carboplatin dosing in clinical trials. This

limited consensus demonstrates the need for further comparison of previous models, and a new model, for GFR in patients with cancer.

In Chapter 4 we work to address these areas of clinical need. We will both develop our own new model and compare previously published models with a large multi-centre validation to assess the most accurate and least bias method for estimating GFR in patients with cancer. First, in Chapter 3 we introduce the data we will use and the relevant clinical and statistical methodology that will guide this work.

Chapter 3

Methodology

This chapter introduces the data, which was collated from multiple cancer treatment centres, and the clinical methodology used for these data. We then give detail of some of the statistical methodology used throughout the remainder of this thesis.

3.1 Clinical methodology

This first section focuses on the clinical and data methodology. We start with an explanation of how data were acquired from cancer treatment hospitals or centres, noting the successive nature of data acquisition both pre- and post-publication of the initial CamGFR model. The individual datasets are then described in order to detail the contents and measurement methods used at the hospitals or centres. We conclude this section by evaluating data inclusion or exclusion criteria.

3.1.1 Data acquisition

The majority of data for this thesis were obtained from patients treated at Addenbrooke's Hospital, Cambridge. These patient data were originally acquired from Addenbrooke's Hospital by Dr. Nicola Ainsworth under an audit project. This project was later expanded to acquire more anonymised data, under a quality improvement project. The project was approved by the audit team (Service Evaluation no. is 907; QSIS reg. no. 6907) and by Dr. Hugo Ford

(divisional director). Dr. Tobias Janowitz (Honorary Consultant and CRUK Scientist) was the lead name on the project.

Additional data were obtained from other cancer treatment hospitals or centres. Candidate centres were identified by literature review and through established clinical collaborations. All major oncological centres within the UK were approached. The following centres provided data: University Beatson West of Scotland Cancer Centre, Glasgow; Hospital Southampton NHS Foundation Trust; a combined South Wales hospital group; Manchester University NHS Foundation Trust; Western General Hospital, Edinburgh; and Barts Health NHS Trust, London. Clinicians had oversight over the quality improvement project (QIP), audit or study protocol at their respective centre. The named contact clinician or clinicians for each of these centres were: Dr P. Mark; Dr M. Fehr and Dr R. Cathomas; Dr G. Bertelli; Dr J. Weaver and Dr P. Monaghan; Dr I. Beh; and Dr J. Shamash, respectively. Data from two additional international centres were acquired: Sahlgrenska University Hospital, Gothenburg; and the Peter MacCallum Cancer Institute, Melbourne. The named contact clinicians were Dr D. Giglio and Dr M Dooley, respectively. Data were fully anonymised prior to transfer for analysis. No identifiable patient information was exchanged in the process of data compilation. Ethical approval for data acquired at the other cancer centres was overseen by the lead clinician from that centre.

The data described above were acquired in succession, with the content and structure of Chapter 4 reflecting this. Originally, a dataset, hereafter referred to as *Cambridge Original*, was acquired from Addenbrooke's Hospital over 6.5-year period from August 2006 until January 2013. A subset of these data were previously used in a publication [2]. Additionally, a second previously published dataset, referred to as *Glasgow* and composed of male patients only with stage I seminoma, was obtained from the Beatson West of Scotland Cancer Centre, Glasgow, United Kingdom [118]. These data informed the analysis in Section 4.1 and the publication of the CamGFR model [60].

After publication of the CamGFR model eight new datasets were acquired. One of these was a second dataset from Addenbrooke's Hospital and is referred to as *Cambridge New*. This dataset consisted of patients treated between January 2013 and April 2019. When both datasets (*Cambridge Original* and *Cambridge New*) are combined for analysis, the data are referred to as *Cambridge*. The other seven new datasets were of patients treated at different cancer centres across the UK and internationally: University Hospital Southampton NHS Foundation Trust [15]; a combined South Wales hospital dataset [106, 90]; Manchester University NHS Foundation Trust; Western General Hospital, Edinburgh; Sahlgrenska University Hospital,

Gothenburg [43]; the Peter MacCallum Cancer Institute, Melbourne [100]; and Barts Health NHS Trust, London, with citations indicating previous publications containing these data. These datasets shall be referred to as *Southampton*, *South Wales*, *Manchester*, *Edinburgh*, *Gothenburg*, *Melbourne*, and *London-Barts* respectively. These data were from a variety of different sources. Some data were historical and had been used in previous publications, which are given above, while others were collected specifically for this thesis.

3.1.2 Content and measurement methods for datasets

All datasets contained the following minimum set of variables for each patient: nuclear medicine measured GFR (mL/min), blood serum creatinine (mg/dL), sex, age (years), height (cm) and weight (kg). Additionally, a final variable for body surface area (m^2) was calculated for all samples. This variable was not measured but calculated using patient height and weight in the DeBois equation [33]:

$$BSA = 0.007184 \times \text{Height}^{0.725} \times \text{Weight}^{0.425}$$

The typical procedure for calculating nuclear medicine GFR (nmGFR) is by an intravenous bolus injection of an exogenous filtration marker, with clearance computed from the amount of the marker administered divided by the area under the curve of plasma concentration over time. The exogenous filtration markers ^{51}Cr-EDTA or ^{99}Tc-DPTA were used, as indicated specifically in Table 3.1.

For nmGFR performed at Addenbrooke's Hospital, Cambridge, a fixed dose of 2 megabecquerel (MBq) was injected and measured at three time points. The protocol for time points was 1, 2 and 4 hours post-injection by venous extraction and direct blood or plasma scintigraphy. For measurements performed at Western General, Edinburgh, where ^{99}Tc-DPTA was used, a dose of 10MBq was administered and measurements were performed 1, 2 and 3 hours post-injection.

Blood serum creatinine (SCr) was measured using either a Jaffe or enzymatic method, both of which are described in Section 2.4.2, by using clinical analysers from a variety of different providers. Information on the clinical analysers and the method used for each centre is provided in Table 3.1.

Data set	Creatinine measurement method	Nuclear tracer
Cambridge Original	Dimension RXL composted Jaffe, non-IDMS-traceable	^{51}Cr-EDTA
Glasgow	An Abbott Architect assay using the Jaffe method, IDMS-traceable	^{51}Cr-EDTA
Southampton	Jaffe method; given the date range, most if not all, non-IDMS-traceable	^{51}Cr-EDTA or ^{99}Tc-DPTA
South Wales	Different centres used either Jaffe or enzymatic methods; IDMS usage not known	^{51}Cr-EDTA
Manchester	Before 08/06/16: Jaffe method, using O'Leary reagent After 08/06/16: Kinetic Jaffe on Siemens ADVIA analyser, IDMS-traceable	^{51}Cr-EDTA
London-Barts	Unknown method; IDMS usage not known	^{51}Cr-EDTA
Edinburgh	An Abbott Architect assay using the Jaffe method non-IDMS-traceable	^{99}Tc-DPTA
Gothenburg	Roche enzymatic colorimetric test, IDMS-traceable	^{51}Cr-EDTA
Melbourne	Jaffe method; non-IDMS-traceable	^{99}Tc-DPTA
Cambridge New	Before 12/12/13: Dimension RXL composite Jaffe, non-IDMS-traceable 12/12/13 - 01/03/16: Siemens ADCEA 2410 Jaffe, IDMS-traceable After 01/03/16: Siemens ADCEA 2410 Enzymatic, IDMS-traceable	^{51}Cr-EDTA

Table 3.1 Summary of the creatinine and GFR measurement methods used at each centre which contributed patient data. IDMS-traceable indicates that the creatinine values were traceable to isotope dilution mass spectrometry standard reference material.

In addition to the minimum set of variables, some of the datasets had additional variables and information:

- Disease diagnosis was known for all patients apart from those from *Melbourne*.
- Specific date differences between GFR and creatinine measurements were known for all, apart from *Southampton*, *South Wales*, and *Melbourne*.
- Ethnicity was known for the majority of the patients from *Cambridge New* and *Manchester*.
- The *Gothenburg* dataset included the additional variables: previous and current treatment, diabetes, insulin, metformin and hypertension.
- The *Glasgow* dataset included the additional variables: serum urea nitrate (mmol/L) and albumin (g/L) concentrations, information on clinical diagnoses such as type 2 diabetes mellitus and hypertension, and information on chemotherapeutic treatment in the form of the area-under-the-curve (AUC) for the administered carboplatin dose.
- The *Cambridge New* and *London-Barts* datasets included several routine biochemical and haematological blood test results. *Cambridge New* included all blood tests that the patient received within 30 days of their nmGFR measurement, whilst *London-Barts*

included only a subset of these variables. These variables included, but are not limited to, albumin (g/L), sodium (mmol/L), serum urea nitrate (mmol/L), alkaline phosphates (U/L), potassium, hemoglobin (g/L), white blood cells count (10^9/L), mean corpuscular volume (fL).

Some of the datasets contained patients who had more than one nmGFR measurement. In particular *Edinburgh*, *Gothenburg*, *Cambridge Original*, and *Cambridge New* were known to have some patients with two or more nmGFR measurements, although this was not known for *Cambridge Original* when the data were first acquired. Additionally, there is a small overlap where some patients have nmGFR measurements in both the *Cambridge Original* and *Cambridge New* datasets.

3.1.3 Inclusion and exclusion criteria

After receiving data from different cancer centres, the data were evaluated and quality controlled. This process involved adjusting units of measurements to be consistent across datasets, correcting obvious data and annotation errors, and excluding data where the reason for suspected errors could not be identified.

Records with data outside the inclusion criteria were removed prior to analysis. The inclusion criteria were:

- Serum creatinine value between 18 µmol/L (0.2 mg/dL) and 400 µmol/L (4.5 mg/dL)
- Age of 18 years or older
- Creatinine and GFR measurements performed within 30 days of each other

The creatinine cut-off values were chosen for the following reasons: the lower value is the typical detection threshold on the measurement assay; the upper value is 3-4 times the upper limit of the normal range of creatinine defined at most centres, thus corresponding to a value at which most clinicians would consider kidney function severely impaired. Only adult patients were included as adolescent physiology can differ considerably from adult physiology. The choice of 30 days was made to minimise errors caused by changing creatinine. This criterion is further assessed in Section 4.3.2. Criteria for the inclusion of patients with repeat GFR measurements changed for the different workflows and these are given in the relevant sections.

3.2 Statistical methodology

This section provides background to the main regression methods used in this thesis. We begin by introducing linear models, and their generalisation to generalised linear models, which themselves can be further generalised to generalised additive models. Then we discuss various methods used for selecting variables for different regression models and the statistics used to compare different models.

3.2.1 Background to linear, generalised linear and generalised additive models

A normal linear regression model assumes that the regression function $E(Y|X)$ for the variable Y is a linear combination of the covariates $X = (X_1, X_2, \ldots, X_p)$. The model can be split into three components:

1. The structure of covariates as a linear predictor

$$\eta = \beta^T X$$

   where X represents the covariates and β are the regression coefficients associated with X
2. The relation $\eta = \mu$ between the expectation $\mu = E(Y|X)$ and the linear predictor η
3. The probability distribution for Y, which is assumed to be normal

$$Y \sim N(\mu, \sigma^2)$$

A generalised linear model (GLM) is a generalisation of a linear model which extends components 2 and 3. Specifically,

2. The linear predictor η is mapped to the expectation $\mu = E(Y|X)$ by the response function h such that

$$\mu = E[Y|X] = h(\eta) = h(\beta^T X)$$

   Equivalently,

$$g(\mu) = \beta^T X$$

where $g = h^{-1}$ is known as the link function. The link function is used to transform between the range of the response variable and range of the linear predictor.

3. The probability distribution for Y is from the exponential family of distributions, where

$$\begin{aligned} P(Y|X,\beta) &= P(Y|\theta(X,\beta),\phi(X,\beta)) \\ &= \exp\left(\frac{y\theta - b(\theta)}{\phi} + c(y,\phi)\right) \end{aligned}$$

where b and c are fixed real-valued functions, θ is the natural parameter, ϕ is the dispersion parameter of an exponential family of distributions. Most common distributions, including the normal, Bernoulli, Poisson, and gamma can be written in this form.

A generalised additive model further generalises component 1, allowing a more general relationship between the linear predictor and the covariates. Specifically

1. Each covariate x_i is related to the linear predictor η by a function f_i

$$\eta = \beta_0 + \sum_{i=1}^{p} f_i(X_i)$$

where the functions f_i may be parametric, non-parametric or semi-parametric.

Parameters are most commonly estimated using maximum likelihood. The likelihood equations are almost always non-linear and cannot be solved analytically. Instead, these equations are often solved using iterative weighted least squares, which is an example of Fisher scoring [114]. In this thesis parameters were evaluated using the R package **stats** [107] for linear and generalised linear models, while the R package **mgcv** [134] was used for generalised additive models.

Box-Cox transformation

Box and Cox produced a method in 1964 that aims to find a power transform for the response variable y in a linear regression such that all the model assumptions are satisfied [10]. Suppose one seeks some power transformation of a positive response y of the form:

$$y^{(\lambda)} = \begin{cases} \frac{y^{\lambda}-1}{\lambda} & \lambda \neq 0 \\ \log(y) & \lambda = 0 \end{cases}$$

such that $y^{(\lambda)}$ satisfies the standard linear model assumptions (including that $y^{(\lambda)}$ has a normal distribution). The Box-Cox method then chooses λ via a likelihood argument. Omitting the intermediate steps, the profile log likelihood for λ, dropping any terms that do not depend on λ, is:

$$L_p(\lambda) = L\left(\hat{\beta}_\lambda, \hat{\sigma}^2_\lambda, \lambda\right) = -\frac{n}{2}\log(RSS_\lambda) = (\lambda+1)\sum \log y_i$$

where RSS_λ is the residual sum of squares for the regression, and $\hat{\beta}_\lambda$ and $\hat{\sigma}^2_\lambda$ are the usual maximum likelihood estimates. A plot of $L_p(\lambda)$ can then summarise the information concerning λ, with an approximate $1-\alpha$ confidence interval given by the λ values satisfying:

$$L_p(\lambda) \geq L_p(\hat{\lambda}) - \frac{1}{2}\chi^2_{1,\alpha}$$

The R package **MASS** [126] was used to calculate these Box-Cox transformations.

3.2.2 Variable selection and regularisation

Variable selection is intended to choose the best subset of predictors which together explain the variability observed in the response variable. This is carried out for a number of reasons: to explain the data in the simplest way (the principle of Occam's Razor states that among several plausible explanations of a phenomenon, the simplest is the best); to avoid the use of unnecessary predictors will add noise when using the model for estimation; and cost – if the model is used for prediction we should avoid including additional predictors which may cost more to measure.

When considering variables to include in a model it is important to respect the hierarchy. This means that an interaction term x_1x_2 between variables x_1 and x_2 should not be included in a model unless both x_1 and x_2 variables are also included. To do so would mean that interpretation of variables would be scale dependent.

There are many different methods to use for variable selection. Below we will outline the methods chosen for the regression models considered in this thesis.

Subset selection

Subset selection aims to choose a set of variables which minimise a given information criterion (IC). Ideally, a model would be fitted for all possible subsets of the p predictors and the information criterion calculated for each model. The model which minimises the IC would then be chosen. However, if we have p predictors, the number of models to be examined is 2^p (not including possible pairwise interactions or different transformations to the predictors), which increases exponentially as p increases. Hence, for computational reasons stepwise methods are commonly used [51, Chapter 3]. These methods work by moving through the hierarchical model space and during each iteration of the algorithm, comparing models using an information criterion. Forward selection, backwards elimination or bidirectional elimination methods can be used. This thesis used bidirectional elimination, where the algorithm works by:

1. Beginning at a given model, such as null model, which only contains the coefficient for the intercept
2. For each iteration, calculating the IC for all possible models that can be reached by adding or subtracting one model term to the current model (ensuring the model remains hierarchical)
3. Choosing to move to the model that has the smallest IC value
4. Repeating the process until no more model terms are added or subtracted

Forward selection and backwards elimination work similarly, except that model terms may only be added during forward selection or removed during backwards elimination.

There are numerous choices for the information criterion used to select the best model, with common choices including Mallows's C_p, Akaike's information criterion (AIC), Bayesian information criterion (BIC) or adjusted R^2. In this thesis we will use BIC, as it tends to select simpler models compared with the other criteria. BIC is defined as:

$$\text{BIC} = -2\log\left(\hat{L}\right) + \log(n)p \tag{3.1}$$

where $\hat{L}$ is the maximised value of the log likelihood, n is the number of observations and p is the number of parameters [113].

In addition to using BIC as an information criterion we consider two cross-validation based criteria [51, Chapter 7] [139, 12, 117]. These have the advantage that they provide a direct estimate of the test error, and help to avoid overfitting, which can be an issue with the above methods. For cross-validation, observations are randomly divided into a training dataset and testing dataset. Parameters of the model are then estimated on the training set, and are then used to estimate the dependent variable in the testing set. The accuracy of these predictions is then summarised, most commonly using root-mean-square-error (RMSE). In k-fold cross-validation, the data are split into k groups. Each of these groups is subsequently used as a test set, while the rest of the groups are combined as a training set. Thus we obtain k estimates of RMSE which are combined by taking the average. For this thesis two values of k are used, 5-fold cross validation, and leave-out-1 cross validation (where $k = n$, the number of observations). When performing 5-fold cross-validation, the statistic has a random component associated with it due to the random nature of splitting the dataset into five groups. To mitigate this randomness the cross-validation is repeated R times, and the cross validation statistics averaged. When using a five-fold cross validation criterion, where the data are randomly split into five subsets, the procedure was repeated 2,000 times with the most commonly observed model selected.

Shrinkage methods

As an alternative to the subset selection methods, techniques that shrink coefficient estimates towards zero can be used. Shrinkage methods have the advantage of subset regression in that they do not suffer from the high variability caused by having a variable either discarded or retained [51, Chapter 3]. Here we discuss elastic net regularisation, which is a generalisation of the lasso and ridge regression methods.

For lasso (least absolute selection and shrinkage operator) regression where the coefficients $\hat{\beta}^{\text{lasso}}$ minimise:

$$(\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta) \text{ such that } \sum_{j=1}^{p} |\beta_j| \leq \lambda \tag{3.2}$$

where $\lambda \geq 0$ is a penalisation parameter which controls the shrinkage, with smaller values increasing the shrinkage of coefficients towards zero. This penalty term results in shrinking the coefficients towards zero by a fixed amount which is dependent on λ. Some coefficients will be exactly zero with a sufficiently small λ, thus lasso is also performing variable selection. The coefficients are calculated using the least angle regression algorithm.

The other common alternative to lasso regression is ridge regression, which uses an alternative penalty term. Here the coefficients $\hat{\beta}^{\text{ridge}}$ minimise:

$$(\mathbf{y}-\mathbf{X}\beta)^T(\mathbf{y}-\mathbf{X}\beta) \text{ such that } \sum_{j=1}^{p}\beta_j^2 \leq \lambda \tag{3.3}$$

With this penalty term the shrinkage of the coefficients is dependent on the value of the coefficients and hence no coefficients will exactly equal zero. Both these penalty terms are combined for elastic net regularization where coefficients $\hat{\beta}^{\text{elastic}}$ minimise:

$$(\mathbf{y}-\mathbf{X}\beta)^T(\mathbf{y}-\mathbf{X}\beta) \text{ such that } \sum_{j=1}^{p}((1-\alpha)|\beta_j|+\alpha\beta_j^2) \leq \lambda \tag{3.4}$$

where $\alpha \in [0,1]$ determines the mix of the two penalties.

To choose the appropriate penalisation parameter λ a cross-validation approach is used, whereby a model is fitted over a grid of λ values. For each model the mean-squared error is calculated using 10-fold cross validation. Finally, the model is refitted using all the available observations and the selected value of the tuning parameter. When choosing the final λ value, it is further recommended to choose the λ which is one standard deviation larger than the minimum value. This is recommended so that the simplest model is chosen, whose accuracy is comparable with the best model [51]. The R package **glmnet** [40] was used to fit all elastic net models.

Variable selection for GAMs

When considering GAM models we used the R package **mgcv** [135] to select smoothness parameters and model variables. We chose to use the semi-parametric P-spline (or penalised B-spline) for any continuous variable [37]. P-splines combine a B-spline basis, with a penalty on the basis coefficients. This penalty term is included to impose smoothness and reduce overfitting.

To estimate the smoothing parameters from the data we used restricted maximum likelihood (REML), which works by treating the smoothing parameter as a variance and iteratively updating it by choosing the value which maximises the restricted likelihood function [133]. However, P-splines and smoothness selection using REML can not remove a smoothed variable from the model entirely. This is due to the P-splines considering some set of non-zero functions

as completely smooth, with further penalisation not changing the fitted function. To overcome this an additional penalty is added, which penalises only functions in the null space of the P-spline penalty [136]. Further detail can be found in the **mgcv** R package [135] documentation and in the provided citations.

3.2.3 Statistics for model comparison

When evaluating or comparing the performance of statistical estimation methods we will discuss three different measures of performance: bias, precision and accuracy. The bias of an estimator is the difference between the expected value of the estimator and the true value of the quantity being estimated. When this difference is zero we call the estimator unbiased. The precision of an estimator is a measure of how similar multiple estimates are to each other, and not necessarily to the true value. In other words, precision is the reciprocal of the variance of an estimator. Bias and precision are then two components of accuracy, which is a measure of closeness of the estimated value and the true value of the quantity.

To assess these three measures four different metrics were used. The median residual was used to assess bias, the residual interquartile range (IQR) was used to assess precision, and both root-mean-squared error (RMSE) and the proportion of patients with an absolute error greater than 20% (P20) were used to assess accuracy. The two measures of accuracy represent a traditional measure of accuracy (RMSE) and a more clinical robust measure of accuracy (P20), where 20% reflects a clinically significant error. For Section 4.1 we additionally used median percentage error (MPE) and absolute percentage error (APE) to assess a models bias and accuracy respectively.

For each of these statistics, 95% confidence intervals were calculated using the bootstrap resampling procedure. Specifically, 2,000 resamples with replacement (where the sample size was the same as the number of data points) were taken from the data. The metric was then calculated for each of these 2,000 samples and using the normal approximation a confidence interval was constructed [25].

To calculate p-values for the hypothesis test between two models for a given metric, another bootstrap procedure was used [36]. This procedure has two samples $\boldsymbol{z}$ and $\boldsymbol{y}$ of lengths n and m respectively, from the probability distributions F and G. In our case F and G were the probability distributions of the residuals for the two equations of interest. The null hypothesis

$H_0 : F = G$ was tested against the alternative hypothesis $H_1 : F \neq G$. Then let $\boldsymbol{x} = [\boldsymbol{z}, \boldsymbol{y}]$, and choose a test statistic $t(\boldsymbol{x})$, which has the form $t(\boldsymbol{x}) = f(\boldsymbol{z}) - f(\boldsymbol{y})$, where the function $f(\cdot)$ was either RMSE, median, IQR or P20. An approximate p-value is calculated by using the following procedure:

- For $r = 1, ..., R$, let $\boldsymbol{x}_r^*$ be a sample with replacement of size $n+m$ from the pooled vector $\boldsymbol{x} = [\boldsymbol{z}, \boldsymbol{y}]$. Let $\boldsymbol{z}_r^*$ be the first n observations of $\boldsymbol{x}_r^*$ and $\boldsymbol{y}_r^*$ the remaining m observations.
- Let
$$t_r^* = t(\boldsymbol{x}_r^*) = f(\boldsymbol{z}_r^*) - f(\boldsymbol{y}_r^*), \quad r = 1, 2, ..., R$$
- Calculate the approximate p-value for the hypothesis test as:
$$\widehat{p} = \frac{1 + \sum_{r=1}^{R} \mathbb{1}_{|t_r^*| \geq |t_{obs}|}}{R+1}$$
where $t_{obs} = t(\boldsymbol{x}) = f(\boldsymbol{z}) - f(\boldsymbol{y})$ is the observed value of the statistic and $\mathbb{1}_A$ is the indicator function which returns 1 if A is true and 0 otherwise. The absolute values of t are considered so as to perform a two-sided test.

It should be noted that the hypothesis test is strictly testing the null hypothesis $H_0 : F = G$ and not the null hypothesis $H_0 : f(F) = f(G)$. When the approximate p-values are calculated with the statistic P20, $\boldsymbol{z}$ and $\boldsymbol{y}$ would be the vectors of percentage differences between the fitted and measured dose values.

Chapter 4

A new creatinine-based model for GFR

This chapter describes the process of developing a new serum creatinine based model to estimate GFR. In Section 4.1 we describe the initial development of a new model, CamGFR. Following this, and after additional data were collected, Section 4.2 validates the CamGFR model and compares it with previously developed models. This section further examines the impact of different methodologies to measure serum creatinine on model accuracy. Given these results, the CamGFR model was redeveloped using a larger dataset in Section 4.3. Finally, some closing remarks and discussion are provided in Section 4.3.3.

4.1 CamGFR: developing the original model

In this section we describe the original development of the CamGFR model [60]. As described in Section 3.1, the data used for this analysis consisted of a dataset of patients with cancer from Addenbrooke's Hospital, *Cambridge Original*, along with a small supplementary dataset of patients all with stage I seminoma from Beatson West of Scotland Cancer Centre, Glasgow. The model was developed and internally validated using the *Cambridge Original* dataset with external validation performed using the *Glasgow* dataset.

Whilst developing the CamGFR model (which ultimately modelled square-root transformed GFR as an ordinary linear model), linear, generalised linear and generalised additive models were all considered, as discussed in Section 3.2.1. The analysis of linear and generalised linear models will be discussed and compared in the following sections. Generalised additive models

were considered but ultimately discounted due the complexity in providing a written equation for a clinician to use and unreliable prediction intervals. However, generalised additive models are later used in Chapter 5 when we consider including additional variables into a model.

Previous models, such as the CKD-EPI or MDRD, modelled GFR using body surface area adjusted units. As discussed in Section 2.2.1, this is due to healthy kidney function being viewed as a function of the patient's BSA and hence general clinical decisions are made using these adjusted units. However, when prescribing a renally excreted drug based on the patient's renal function, such as carboplatin, GFR in unadjusted units is used, which can also be referred to as absolute GFR. Furthermore, there is no universal consensus on which equation to use to calculate a patient's body surface area, with the DuBois & DuBois [33] and Haycock [52], two commonly used equations. This lack of consensus and the small differences between alternative equations for GFR can result in additional errors being introduced. Thus to mitigate for this we chose to model GFR in its natural units, mL/min, and considered BSA (calculated using the DuBois & DuBois equation) as a variable for the model.

4.1.1 Data cleaning and description

Before analyses were performed, data were examined and cleaned. This involved first graphically examining the data and editing any obvious mistakes. For example, one patient was identified as having their height and weight interchanged. If the source of the mistake was not obvious then those data were excluded. Data were also excluded if they did not meet the inclusion criteria described in Section 3.1.3; 193 were removed as the time between GFR and creatinine measurement exceeded 30 days and 19 patients with an age of less than 18 years at the time of the measured GFR date were removed. Due to early anonymisation of the data used for the original analysis, no information was available as to whether any nuclear medicine GFR measurements were from patients who had repeat measurements, hence, data were not excluded based on this criterion.

After the above patients were excluded, there was a total of 2,471 patients in the *Cambridge Original* dataset and 111 in the *Glasgow* dataset. Table 4.1 summarises the patient characteristics for these two datasets.

The patient population in the *Cambridge Original* dataset was diverse as evidenced by a wide range of measured GFR (11–211 mL/min), age (18–92 years), and serum creatinine

	Cambridge original			*Glasgow*		
	median (IQR)	range	mean (SD)	median (IQR)	range	mean (SD)
Age [years]	61 (50-69)	18-92	58 (15)	39 (33-46)	21-69	40 (10)
Weight [kg]	73 (62-85)	37-163	75 (17)	86 (76-98)	51-161	87 (18)
Height [cm]	168 (161-176)	125-200	168 (10)	178 (174-182)	131-192	177 (8)
BSA [m^2]	1.84 (1.67-2.00)	1.24-2.79	1.84 (0.23)	2.04 (1.92-2.15)	1.48-2.73	2.04 (0.21)
Serum creatinine [mg/dL]	0.91 (0.77-1.07)	0.29-5.62	0.96 (0.34)	0.92 (0.81-1.01)	0.62-1.45	0.92 (0.14)
^{51}Cr-EDTA GFR [mL/min]	81 (63-103)	11-211	84 (29)	113 (101-131)	45-202	116 (26)
^{51}Cr-EDTA GFR [mL/min/1.73m^2]	79 (62-94)	10-201	78 (24)	95 (88-105)	35-160	98 (17)
	No. (%)			No. (%)		
^{51}Cr-EDTA GFR						
<40 mL/min/1.73m^2	131 (5)			1 (1)		
[40 , 60) mL/min/1.73m^2	430 (17)			1 (1)		
[60 , 90) mL/min/1.73m^2	1134 (46)			34 (30)		
≥90 mL/min/1.73m^2	775 (31)			75 (68)		
Sex						
Female	1398 (57)			0 (0)		
Male	1073 (43)			111 (100)		

Abbreviations: BSA - body surface area, GFR - glomerular filtration rate, IQR - inter quartile range
[a,b) indicates values ranging from a to b excluding a but including b
Modified from Table 1 in [60]

Table 4.1 Patient characteristics for the *Cambridge Original* and *Glasgow* datasets.

(0.29–5.62 mg/dL). Patients had a higher mean measured GFR (78 mL/min/1.73m^2) than previously published studies that were mostly reliant on patients with kidney disease rather than with cancer, as discussed in Section 2.4.3. In particular, it was significantly higher than the patient population used to develop CKD-EPI (68 mL/min/1.73m^2) and 186-MDRD (39.8 mL/min/1.73m^2) models. Furthermore, the median (79 mL/min/1.73m^2) was higher than the median of the patient population used to develop the Wright model (73 mL/min/1.73m^2).

The patient population contained a higher proportion of males (57%) compared with females (43%). The median age of the patients was 61.1 years, which is older than the UK population average but was expected given that all of the patients have had cancer. In fact, the average age of a cancer patient at diagnosis was 68 in the UK between 2014 and 2016 [14]. Male patients in the dataset were younger on average than female patients (median 56 years compared with 64 years). This difference can be explained by the patients' type of cancer which is treated using carboplatin chemotherapy, in particular germ cell tumours and ovarian cancer. Of the 1,250 patients with a categorised diagnosis, 231 males were categorised with a germ cell tumour and 404 females were categorised with a gynaecological cancer, which were the largest cancer subgroups for each sex, and these patients had a median age of 35 and 65 respectively.

The patients' blood serum creatinine (SCr) ranged from 0.29 to 5.62 mg/dL. However, 90% of the data were distributed between 0.59 and 1.54 mg/dL, with a mean value of 0.96 mg/dL.

Serum creatinine had a skewed distribution with a long tail to the higher values. Males had higher values of serum creatinine, mean 1.06 mg/dL compared with 0.883 mg/dL for females. However, as males tend to have greater muscle mass, and as muscle mass is correlated with serum creatinine, this was not an unexpected finding. The variables height, weight and surface area all had approximately normal distributions.

Figure 4.1 shows the bivariate relationship between GFR and the other variables. Height, weight and BSA have an approximate liner relationship with GFR. For weight and BSA, there appears to be a difference in the slope of the fitted lines between genders, whilst for height there appears to be a difference in the intercept. The relationship between GFR and age is approximately flat before the age of 40 years with a linear decline afterward. This is the same as the observed relationship between GFR and age in healthy kidney donor patients [38]. Any small differences between cancer and kidney donor patients are discussed further in Section 6.3

The relationship of GFR with creatinine is non-linear and heteroscedastic, evidenced by the decreasing variability of GFR with increasing serum creatinine. Furthermore, the plot displayed in Figure 4.1 suggests that a log transformation of creatinine would be beneficial, both to mitigate the skewness in the data and to approximate a more linear relationship between GFR and log(SCr). This is the transformation that most previously developed models have used.

4.1.2 Model development

We used the cleaned and transformed data to develop a new model for GFR in patients with cancer. For this, the *Cambridge Original* dataset was divided randomly into a development and internal validation set at a ratio of 4:1. The new model was subsequently developed using the development dataset and internally validated against previously published models using the separate internal validation dataset.

Distribution of GFR

The first step to fit a generalised linear model is to decide on the distribution of the error terms. By definition, the minimum for GFR is zero in the absence of functional glomeruli and the values for GFR are thus strictly positive with a skew to the right. Common distributions with these attributes are the gamma and log-normal distributions, however the normal distribution

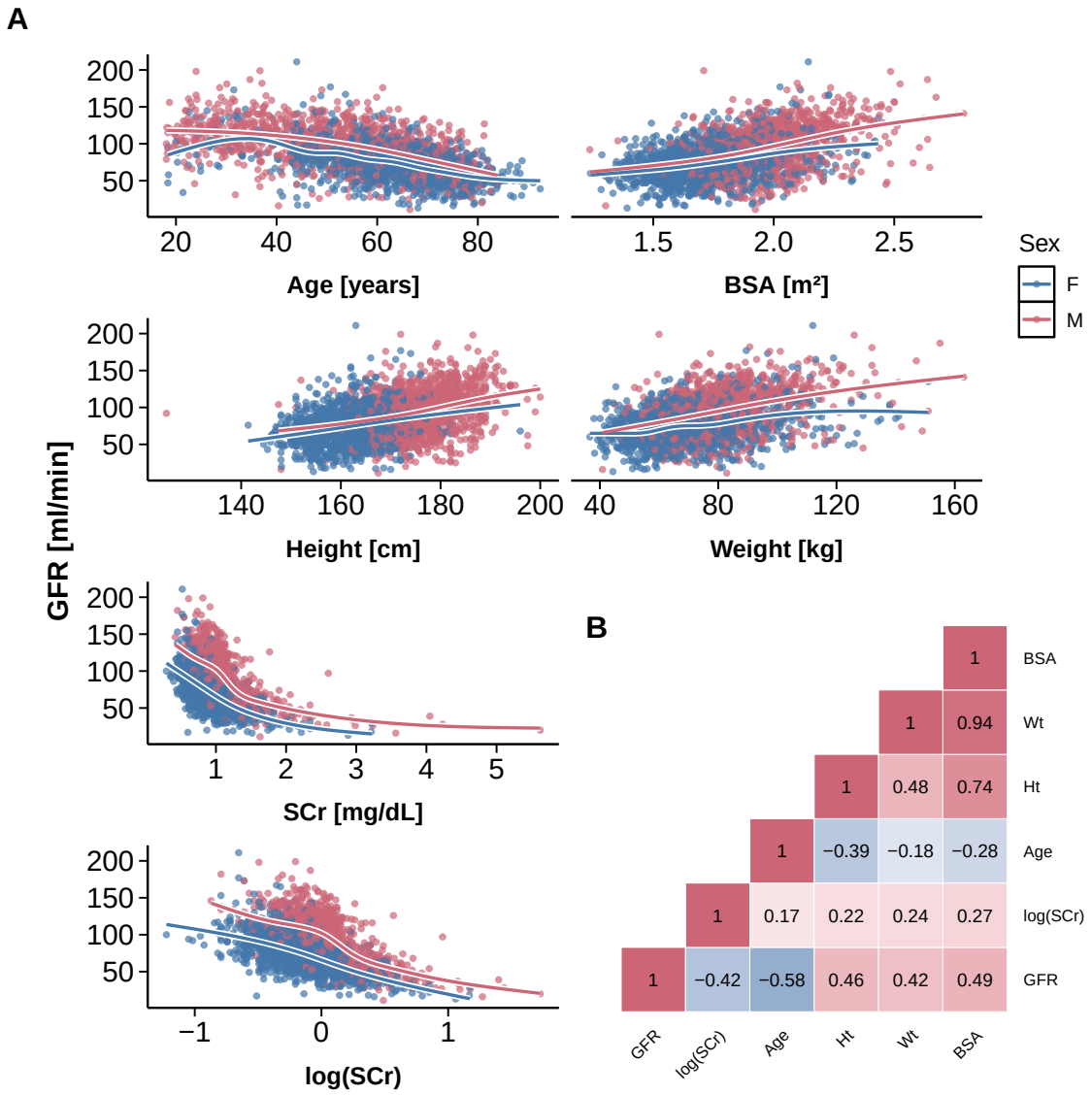


Figure 4.1 Relationship between GFR and the continuous explanatory variable for patients in the *Cambridge Original* dataset. (A) Plot of the GFR against each of the other continuous variables (age, BSA, height, weight, SCr) and the natural logarithm of serum creatinine (log(SCr)) for the *Cambridge Original* dataset. Smoothed trend lines are calculated using a generalised additive model with a cubic spline basis. Points and lines are coloured by sex. (B) Pearson correlation matrix for the continuous variables in the *Cambridge Original* dataset. Abbreviations: BSA - body surface area, SCr - serum creatinine, GFR - glomerular filtration rate.

may also be suitable. We plotted the data to examine the best fit and from Figure 4.2 (A) and (B) it is evident that the gamma distribution best describes the data distribution, demonstrated

by the theoretical densities and quantities matching the observed density and quantities in the histogram and quantile-quantile plot respectively.

For a GLM, not only is there a choice for the distribution of the response variable, but there is also a choice for the link function. A common option is to choose the natural or canonical link because it has the small advantage of simplifying some calculations. The canonical link for the gamma distribution is the inverse function. However, this raises some issues as it would be possible, though unlikely, for the model to predict negative fitted values for GFR, which is paradoxical and not desirable. For the gamma distribution, another more common choice for the link function is the log function, which was chosen as it cannot produce negative fitted values.

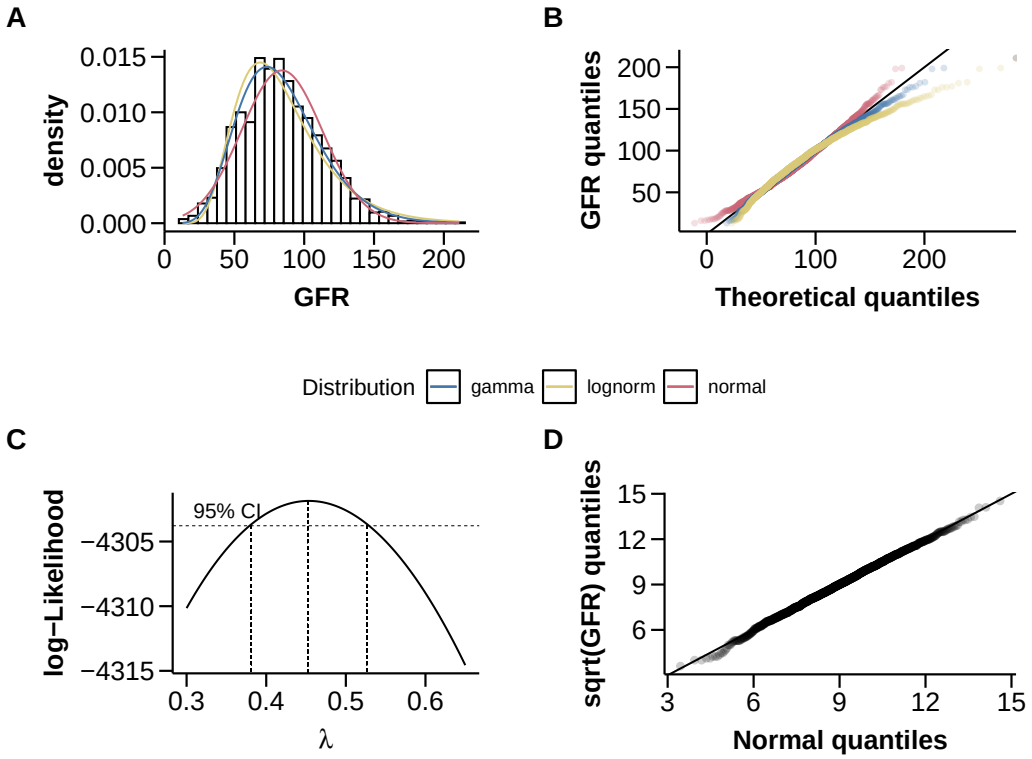


Figure 4.2 Fitted distributions for GFR along with the Box-Cox log-likelihood showing square-root transformation is suitable. (A) Histogram of GFR along with the probability density curve for a gamma, log-normal and normal distribution. The parameters for each distribution are calculated via maximum likelihood. (B) Quantile-quantile plot comparing the gamma, log-normal and normal distribution. (C) Profile log-likelihood for Box-Cox method. Horizontal lines are drawn for the maximum likelihood and the lower and upper values for the 95% confidence interval. (D) Normal distribution quantile-quantile plot for square-root transformed GFR.

An alternative to fitting a GLM would be to transform the GFR to be normally distributed and fit a linear model. To find the most suitable transformation the Box-Cox method was used, which calculates the power transformation to apply to the data to produce approximately normally distributed residuals [10]. This method requires an initial model that includes all key predictor variables used in previously published creatinine based models for GFR (creatinine, age, sex, body surface area, height, and weight).

$$GFR = \beta_0 + \beta_1 \text{SCr} + \beta_2 \text{Age} + \beta_3 \text{Wt} + \beta_4 \text{Ht} + \beta_5 \text{BSA} + \beta_6 \text{Sex}_\text{M} + \varepsilon \tag{4.1}$$

where ε is normally distributed with mean 0 and a given variance. In this equation and following equations variable units and definition are as defined in Table 2.1 and additionally Sex_M equals 1 for males and 0 for females. The Box-Cox method produced the maximum likelihood estimate for the power transformation of $\lambda = 0.473$ (95% CI 0.400 - 0.547), Figure 4.2 (C). This confidence interval contained 0.5, corresponding to the square root transformation. A quantile-quantile plot (Figure 4.2 D) confirms that square-root transformed GFR is approximately normally distributed. Furthermore, this transformation guarantees that predicted GFR values cannot be negative. Most previous models have used a log-transformation for GFR, however, since the Box-Cox confidence interval for λ does not include zero, there is little support for this being an appropriate transformation.

Variable selection

The next stage in fitting the model was to select the variables that are to be included. As described in Section 4.1.2, several different approaches were used to determine which variables should be considered. For this analysis we chose to use a stepwise selection procedure with three different criteria, five-fold cross validation, leave-one-out cross-validation and Bayesian information criterion (BIC). These selection criteria were used for both the linear model with transformed GFR and the generalised linear model.

Linear model with transformed GFR

An assumption for linear modelling is that the relationship between the independent and every dependent variable is linear. We assessed this graphically using residual against fitted value plots obtained by fitting the initial model (Equation 4.1) with the response variable now equal

to square-root transformed GFR. The plots for age, height, weight and body surface area showed no patterns and hence no transformation was required (Figure 4.3 A-D). In contrast, the plot for creatinine suggested a non-linear relationship between creatinine and square-root GFR (Figure 4.3 E). Through a sequence of transformation attempts and reassessments we determined that natural logarithm transformation as well as inclusion of squared and cubic terms of this transformation, achieved linearity (Figure 4.3 F-H). These additional polynomial terms also significantly improved the initial model containing all variables (p-value <0.0001, F-test).

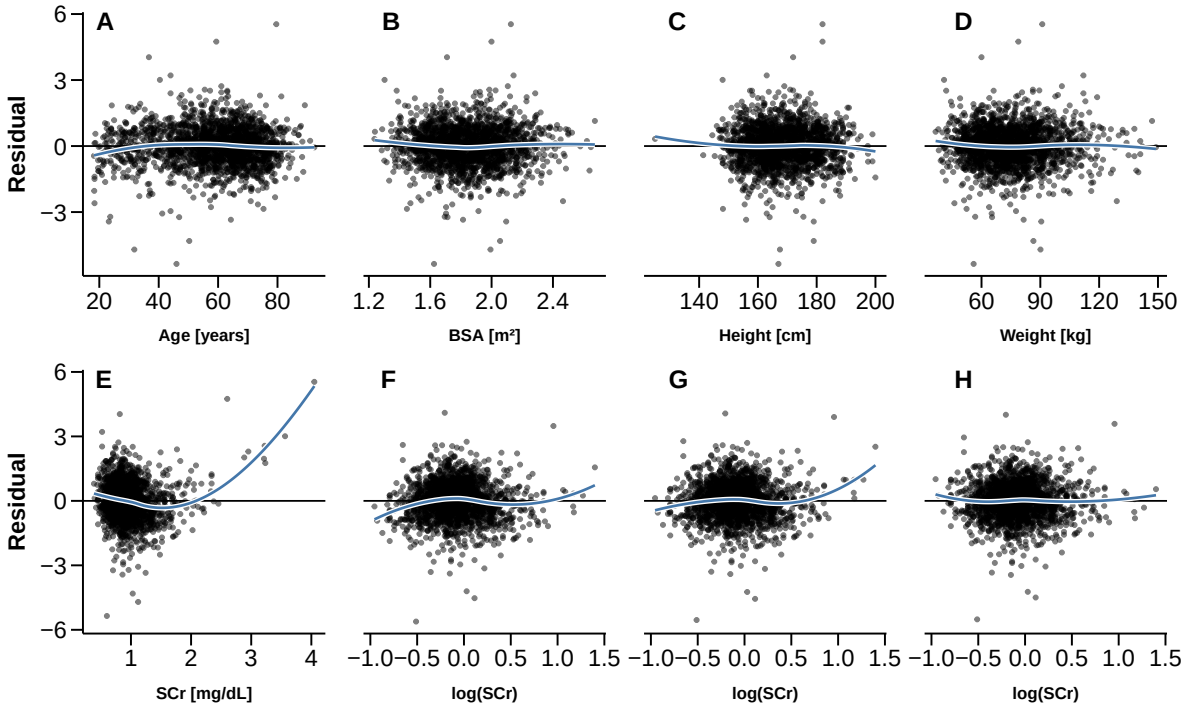


Figure 4.3 Model residuals against explanatory variables for the initial model and after including square and cubic log(SCr) terms. (A-E) Model residuals from the fitted initial model (Equation 4.1) plotted against each of the explanatory variables in that model: age (A), body surface area (B), height (C), weight (D), and creatinine (E). (F-H) Model residual against natural logarithm of creatinine after creatinine is replaced by transformed creatinine in the simple model (F), an additional squared log(SCr) (G) as well as cubic log(SCr) (H) are added to the initial model. All data smoothing lines are loess curves with the span parameter set to 0.9. Figure is adapted from Supplemental Figure 3 in [60]

Next, we determined the variables to include in the final model. Initial exploratory analyses indicated that pairwise interaction terms might lead to improved estimation of GFR. Hence, models with pairwise interaction terms were considered with the exception of squared and cubic log(SCr). It was required that the model remained hierarchical, that is, a model with a pairwise interaction between two variables must also contain both those variables.

After performing stepwise selection with the three different criteria, the same model was selected using the leave-one-out and BIC criteria. The same model was also the most frequently returned model at a count of 854 from the 2,000 repetitions of the five-fold cross-validation criterion, representing a four times more frequent return than the second most frequent model. This model contained the variables age, BSA, sex, log(SCr) along with its squared and cubic terms, and the age-sex and BSA-age interaction terms. The coefficients of the model used for the internal validation (sqrt LM) were calculated using the full validation dataset and the R package **stats**.

Gamma generalised linear model

A similar procedure was used to select the variables for the gamma generalised model. As above, all pairwise interaction terms were considered along with squared and cubic log(SCr) terms. When applying the stepwise selection procedure for gamma GLM the three different criteria selected different models. When using BIC as the criterion, the variables selected differed from the variables selected for the linear model with sqrt-transformed GFR by changing the BSA-age interaction term to a log(SCr)-sex interaction term. The variables selected using leave-one-out and leave-out-n additionally included the height and the log(SCr)-age interaction term. The leave-one-out did not include the cubic log(SCr) term.

The model selected for the leave-out-n criteria is the model that was observed most frequently in the 2,000 repetitions. For the linear model above, the model was observed almost four times more often than the next most commonly selected model. However, for the gamma GLM the top three models were observed 584, 471, and 401 times out of the 2,000 iterations. Given this, and as simpler models are preferable, we chose the model selected using BIC to compare against the square-root transformed GFR linear model. As above, the coefficients of the model used for the internal validation (gamma GLM) were calculated using the full validation dataset and the R package **stats**.

4.1.3 Model validation and assessment

Internal validation

The selected square-root transformed linear model (sqrt LM) and gamma generalised linear model (gamma GLM) from Section 4.1.2 were compared with each other and previously published models using the internal validation dataset. This dataset had no role in developing the models or fitting the parameters. If a previously published model estimated GFR using BSA adjusted units, the estimated value was first unadjusted before being used in the comparison. As discussed in Section 3.2.3, five statistics were used to compare models for this publication. These were median percentage error (MPE), root-mean-squared-error (RMSE), interquartile range of the residuals (residual IQR), and median absolute percentage error (MAPE). All statistics were calculated using the original scale of GFR. These statistics assess different aspects of a model's fit, namely its accuracy, precision, and bias, with the final statistics assessing the number of patients who would have a clinically relevant estimation error.

Table 4.2 provides summary statistics to compare the sqrt LM with the gamma GLM along with other previously published models using the internal validating dataset. The sqrt LM was the most accurate and precise, evidenced by lower RMSE, residual IQR and median APE values. No significant bias was observed for either model. Given these results, we chose to use the sqrt LM model for future work. This was despite the advantage of modelling GFR on the original scale provided by the gamma model.

The sqrt LM was also the most accurate model when compared with previously developed models. It had the lowest RMSE at 15.00 mL/min (95% CI 14.12–16.00), compared with 16.30 mL/min (95% CI 15.34–17.38) for the CKD-EPI and 23.75 mL/min (95% CI 22.36–25.33) for the Cockcroft-Gault.

Most published models were developed from datasets including many patients with impaired kidney function. To reflect this, we performed a subgroup analysis in patients with eGFR less than or more than 60 mL/min. The number of patients with an eGFR less than 60 mL/min varied for each model, ranging from 38 for the Mayo model to 160 for the Jelliffe model, and the Sqrt LM model had 88 patients. The Wright model had the lowest RMSE of 11.53 mL/min (95% CI 9.95–13.70) in the subset of patients with an eGFR of less than 60 mL/min, followed by the new linear model (11.64 mL/min, 95% CI 10.15–13.66). For patients with eGFR more

Model	Median eGFR [mL/min]	RMSE [mL/min]	Median residual [mL/min]	Residual IQR [mL/min]	Median PE [%]	Median APE [%]
Cambridge original validation						
Sqrt LM	79.63	**15.00**	0.51	**17.66**	0.65	**11.59**
Gamma GLM	80.00	15.17	-0.10	18.58	-0.11	12.01
CKD-EPI	80.63	16.30	**-0.03**	20.04	**-0.03**	12.97
MDRD	73.51	17.46	6.99	19.23	8.93	14.60
Cockcroft Gault	79.01	23.75	-0.79	24.47	-1.04	15.55
Jelliffe	71.39	21.39	7.52	21.82	10.90	17.40
Wright	86.14	22.77	-6.12	23.85	-8.55	13.88
Mayo	93.25	23.73	-14.91	22.64	-19.41	20.74
Martin	85.87	25.49	-8.20	24.79	-10.34	16.18
Glasgow						
CamGFR	111.82	**18.94**	**1.61**	**22.92**	**1.39**	**11.18**
CKD-EPI	120.62	21.33	-5.27	27.52	-4.41	12.95
MDRD	104.44	22.88	10.00	26.96	9.22	12.25
Cockcroft Gault	126.52	32.32	-18.04	31.57	-14.14	16.30
Jelliffe	104.97	24.38	9.11	27.69	7.55	13.65
Wright	126.27	26.49	-12.39	29.09	-9.91	14.80
Mayo	146.21	36.50	-31.52	27.26	-27.43	27.56
Martin	133.37	37.74	-23.87	32.83	-22.34	22.34

Abbreviations: eGFR - estimated, RMSE - root-mean-squared-error, IQR - inter quartile range
PE - percentage error, APE - absolute percentage error
SqrtLM - square root transformed linear model fitted on model development dataset
GammaGLM - gamma distribution generalized linear model fitted on model development dataset
CamGFR - final square root transformed linear model fitted on the full Cambridge original dataset
CKD-EPI - Chronic Kidney Disease Epidemiology model, MDRD - Modification of Diet in Renal Disease model

Table 4.2 Summary statistics for the newly developed and previously published models on the internal (*Cambridge Original*) validation and external (*Glasgow*) validation datasets.

than 60 mL/min, the new model was the most accurate and least biased with a RMSE of 15.63 mL/min (95% CI 14.62–16.79) and a median residual of 0.61ml/min.

There was a noticeable difference in all models' accuracy between the two subgroups where patients with a smaller eGFR had more accurate estimated values. This is likely due to the fact that GFR is modelled on a different scale from the scale at which accuracy is calculated. Specifically, for the new linear model, it is modelled on the square root scale where the variance of the residuals is constant. However, when the estimated values are squared, the variability in the larger estimated values is proportionally higher than that of smaller estimated values.

Assessing model robustness

After this comparison we also assessed the dependence of the final model on the random partitioning of the dataset. Hence, the full dataset was partitioned 100 times into different development and internal validation datasets in a 4:1 split. All three model selection methods were repeated on each one of these repartitioned datasets, with 100 rather than 2,000 repetitions for the five-fold cross-validation. Our final model was selected most frequently in the 100 datasets for all three stepwise selection criterion. Using the five-fold cross-validation criterion it was chosen 22 times and the next most common model, which was chosen 20 times, only differed from the final model by an additional height variable. The leave-one-out cross-validation criterion returned the new model 16 times and the second most frequent model was returned 15 times. Again, the second most frequent model only differed by an additional height variable from the new model. The BIC selection criterion selected the new model 31 times with the next most frequently selected 15 times. We found that for 14 random data splits all three methods returned the final model. The next most frequent model, which contained an additional height term, was only selected five times by all three methods.

Once the variables selected were confirmed, the final coefficients of the model were calculated from the full *Cambridge Original* dataset. This model is referred to as the CamGFR model, or the original CamGFR model in later sections. Figure 4.4 provides diagnostic plots for the new model confirming that no single data point was influential in the full dataset (highest Cook's distance value: 0.094). The residual against fitted value plot showed an overall good fit to the data (Figure 4.4 A). Importantly, there was no heteroscedasticity in the final linear model and we thus confirmed that calculation of confidence intervals (or prediction intervals) for the estimated GFR values was appropriate (Figure 4.4 B).

The CamGFR model with coefficients rounded to 3 significant figures was the following

$$\begin{aligned}\sqrt{GFR} = 1.81 + 0.0191\,\mathrm{Age} + 4.73\,\mathrm{BSA} - 3.72\log(\mathrm{SCr}) - 0.914\log(\mathrm{SCr})^2 + \\ 1.063\log(\mathrm{SCr})^3 + (0.0202 + 0.0125\,\mathrm{Age})\,\mathrm{Sex_M} + 0.0297\,\mathrm{Age}\,\mathrm{BSA} + \varepsilon\end{aligned} \tag{4.2}$$

where:

- GFR is glomerular filtration rate with units mL/min
- Age has the units years

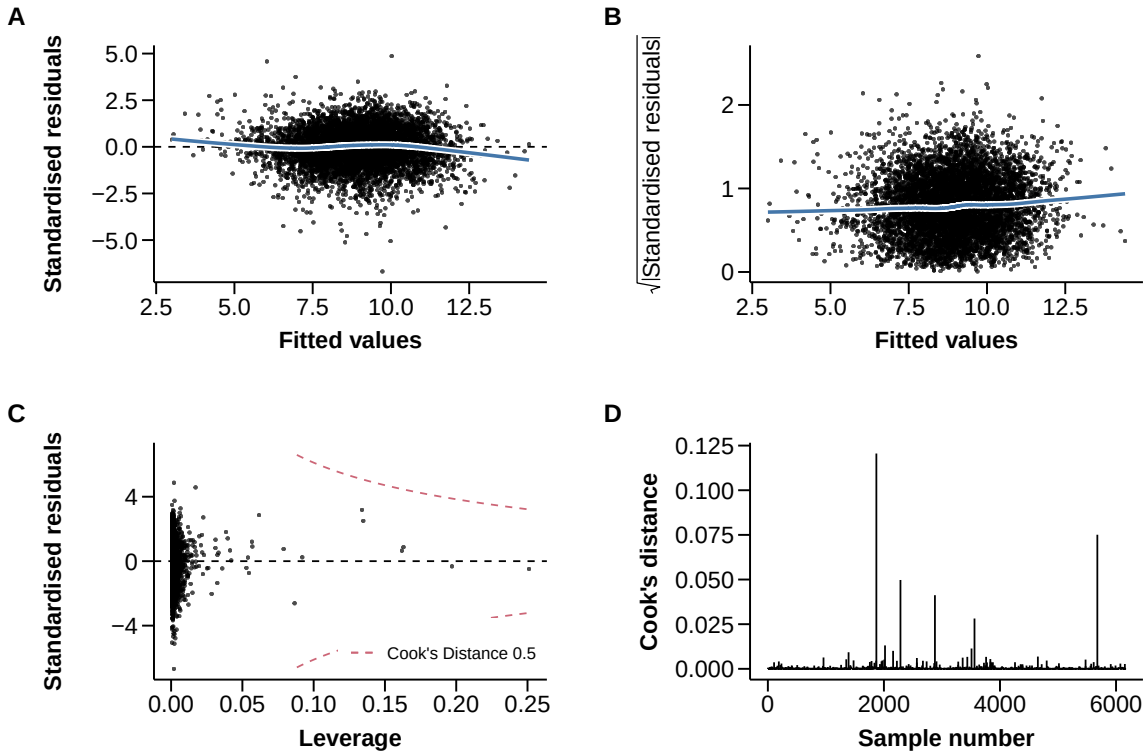


Figure 4.4 Model diagnostic plots for CamGFR fitted on the full *Cambridge Original* dataset. (A) Residual against fitted value plot. (B) Scale-location plot. For both these plots the fitted curve is calculated using loess with a span equal to 0.5. (C) Leverage against standardised residuals where the dashed red line indicates a Cook's distance of 0.5. (D) Bar plot of Cook's distances for all samples used to fit the final model.

- BSA is body surface area with units m^2 calculated using the DuBois & DuBois equation [33]
- SCr is blood serum creatinine concentration with units mg/dL
- Sex_M is equal to 1 when the patient is male and 0 when the patient is female
- ε are independent, mean zero normally distributed random variables with a constant variance

This model had a multiple R^2 value equal to 0.726 and the standard error of the residuals was 0.842. The standard errors of the coefficients can be seen in Table 4.5.

External validation

Finally, we externally validated the model using a dataset from a different cancer centre. Table 4.2 provides the summary statistics for comparing the CamGFR model with the previously published model on the *Glasgow* external validation dataset. As with the internal validation, the CamGFR model remained the most accurate compared with all other published models or BSA adjusted published models. The RMSE for the GFR calculated with the new model was 18.93 mL/min compared with 21.33 mL/min for BSA adjusted CKD-EPI, and 32.32 mL/min for Cockcroft-Gault. Of the 111 patients in the external validation dataset, 105 (94.6%) had the measured GFR within the 95% confidence interval.

Dosing for carboplatin

We investigated the relevance of our findings to chemotherapy treatment in clinical practice by assessing the dose accuracy for the chemotherapy drug carboplatin, which is excreted renally. To calculate the dose for carboplatin the Calvert formula [13] is used:

$$\text{Dose [mg]} = \text{AUC [mg/mL/min]} \left(\text{GFR [mL/min]} + 25\right) \tag{4.3}$$

where AUC is the target area under the curve for the dose concentration time curve. For a given AUC this is linear in GFR, hence any improvement in the accuracy of GFR estimation will translate to an improvement in the accuracy of the dose given as confirmed by our calculations. This improvement is clinically significant because the fraction of patients that receive a dose inaccuracy of more than 20% is reduced by using eGFR calculated using the CamGFR model.

4.1.4 Discussion

This section has described our work to develop a new model to estimate GFR in patients with cancer [60]. There are multiple appropriate approaches with differing advantages and disadvantages that could have been employed. Ultimately, for reasons discussed in this chapter, we chose to model square root transformation of GFR as a linear combination of explanatory variables with a cubic polynomial fitted for the log transformation of serum creatinine to capture the complex relationship between GFR and serum creatinine.

The CamGFR model is distinguished from previous models by the patient population used to develop it. With the exception of the Wright model [137] and the Giglio model [43] previous models to estimate GFR have been developed using data from non-cancer patients. For non-cancer patients a common reason to measure GFR is to diagnose or monitor chronic kidney disease. Hence, these patient populations are often particularly enriched for patients with impaired kidney function compared with both our dataset and the normal population.

One of the primary reasons to measure GFR for patients with cancer is for dose calculations for carboplatin. However, as discussed in Section 2.5 there is little consensus in the choice of method to calculate GFR for this purpose. The CamGFR model was developed in an attempt to rectify this issue. CamGFR reduces the fraction of patients who receive a carboplatin dose that is more than 20% different from the dose calculated using ^{51}Cr-EDTA GFR, compared with previous models.

To allow clinicians to utilise the CamGFR model, we developed a web application using the R package shiny [16], which is a web application framework for R. This application allows a user to input a patient's demographics and serum creatinine data, and returns estimated GFR using the CamGFR and the next best model, CKD-EPI. Importantly, for the CamGFR model the application provides confidence intervals for the predicted value and an estimate of the probability that the predicted value is above or below a given threshold. This gives clinicians or other users a sense of the accuracy of the estimated value, which should be considered when making patient management decisions such as drug prescriptions. No other currently available model provides an estimate of the accuracy of its estimated model. Since the publication of the CamGFR model in July 2017, the application has had an average of 56 uses per week. A screen shot of the application is provided in Figure 4.5.

Of course, the CamGFR model has some limitations. First, using creatinine as the primary explanatory variable in predicting GFR has its own limitations. Other predicting variables such as cystatin C have been proposed as alternative or adjunctive variables. However, cystatin C measurements are not routinely collected and so were not available. Furthermore, use of cystatin C has been called into question in patients with cancer because levels may fluctuate in a cancer-dependent and kidney-function-independent manner [70]. Additionally, we did not analyse measurements of albumin, muscle mass, information on dietary and fluid intake, or co-morbidities (such as diabetes mellitus) due to lack of data.

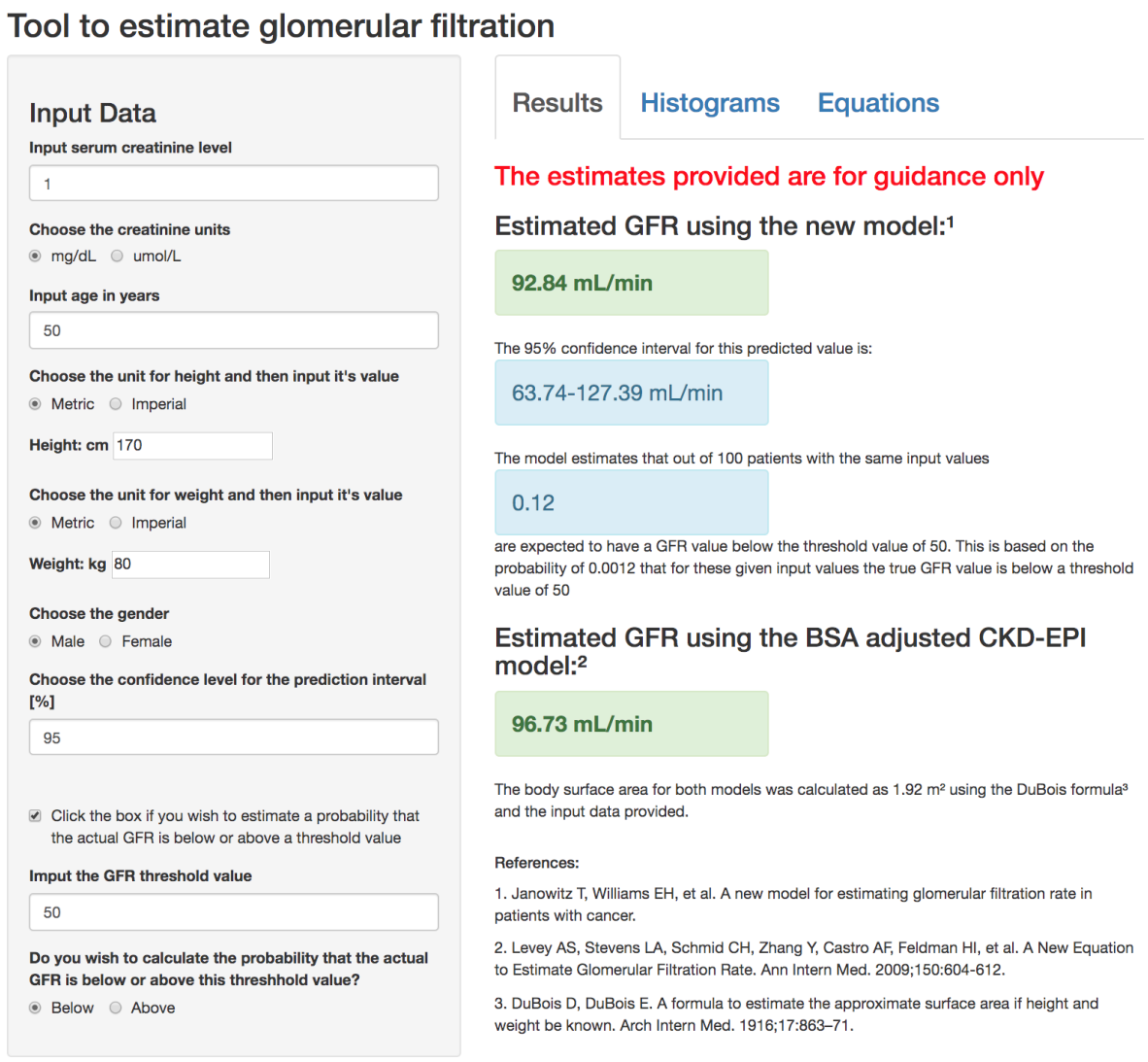


Figure 4.5 Screen shot of the shiny application for the CamGFR model. The application allows users to input a patient's information and it returns an estimate of GFR from the CamGFR model along with a confidence interval for this predicted value and a probability of the estimated value being above or below a chosen threshold. Histograms are provided in another tab so that the user can compare the patient's data to those used to build the CamGFR model.
Link to the application: tavarelab.cruk.cam.ac.uk/JanowitzWilliamsGFR

This study is also limited by the Caucasian-only population, a fact that is attributable to the single centre population demographics. Others have shown that adjustment factors improve GFR prediction for patients who are ethnically black [77, 74]. We were not able to assess whether such an adjustment factor would be required for the CamGFR model. The final coefficients reported in our study may be centre-dependent as a result of differing creatinine results and patient population. As discussed in Section 2.4.2, the issues with centre-dependent

creatinine values is being addressed by international guidelines to standardise creatinine reporting. These guidelines were not yet implemented at Addenbrooke's Hospital at the time of our initial data collection. It is now recommended that serum creatinine values should be calibrated to IDMS-traceable reference samples. This issue is further exacerbated by the small and male-only validation dataset, which does not provide sufficient evidence for generally applicability of the CamGFR model. Further investigation is required to validate or adjust the CamGFR model with data from multiple cancer centres.

4.2 Creatinine data analysis and validation of the CamGFR model

After the publication of the CamGFR model, we sought to collect further data, particularly from different cancer centres, in order to address some of the limitations raised in the previous section. Using these data, we addressed two main points. First, we assessed the differences in the serum creatinine resulting from inter-centre variability, calibration to IDMS reference material, and different assay methodology. We then validated the CamGFR model using a more extensive and diverse patient population.

In total, eight new datasets were acquired, consisting of 6,952 patients and 7,944 nmGFR measurements. The number of nmGFR measurements was greater than the number of patients because a subgroup of patients had multiple nmGFR measurements taken during their cancer management (mean 1.14, range 1-18). Of the 7,944 GFR measurements, 3,919 had creatinine measured that was traceable to an IDMS standard reference material (IDMS-traceable). For a subset of these patients, additional biochemical results such as albumin and urea were available along with creatinine. Ethnicity was known for 2,758 of the patients. However, 95% of patients with known ethnicity were white. Table 4.3 provides a patient population summary for each of the eight new datasets. The number of patients in each dataset ranged from 108 in *London-Barts* to 2,737 in the *Cambridge New* dataset. Specific disease or cancer site diagnosis information was known for the majority of patients from all datasets apart from *Melbourne*. Overall, the majority of creatinine was measured using a non-IDMS-traceable method.

The methods used to measure GFR and serum creatinine are discussed in Section 3.1.2 and detailed in Table 3.1. The methodology used to measure creatinine or GFR differed among the eight new datasets. For *Cambridge New* and *Manchester*, the creatinine methodology

also changed within the date range of the dataset. Specifically, in the *Cambridge New* dataset creatinine was first measured using a non-IDMS-traceable Jaffe assay, then an IDMS-traceable Jaffe assay, and finally an IDMS-traceable enzymatic assay. Similarly, in the *Manchester* dataset creatinine was originally measured using a non-IDMS-traceable Jaffe assay which was changed to an IDMS-traceable Jaffe assay. Overall, the majority of creatinine was measured using a non-IDMS-traceable method. For GFR, two of the datasets (*Edinburgh* and *Melbourne*) used ^{99}Tc-DPTA as the nuclear tracer of the nmGFR measurement, with the others using ^{51}Cr-EDTA.

Dataset	Date range	No. of patients	No. of nmGFR	No. with diagnosis	No. with IDMS	Patient composition
Southampton	01/01/99 - 01/01/12	436	436	436	0	Cancer - Seminoma
Wales	01/01/05 - 01/01/11	157	157	157	0	Cancer - Seminoma and Ovarian
Manchester	02/04/12 - 07/02/17	2,068	2,068	1,591	282	Cancer general
London-Barts	16/08/10 - 12/12/17	108	108	108	0	Cancer - Seminoma
Edinburgh	01/13/11 - 21/12/17	598	628	585	0	Cancer and nephrology
Gothenburg	01/02/08 - 28/08/12	627	914	627	914	Cancer and two non-cancer patients
Melbourne	05/01/00 - 28/01/05	311	311	0	0	Cancer general
Cambridge new	01/02/13 - 20/04/19	2,737	3,452	2,548	2,723	Cancer and non-cancer general

Date range refers to the range of dates within which patients received treatment.
No. of patients refers to the total number of individuals in the respective datasets.
No of nmGFR (or number of rows of the dataset) is the total number of nmGFR measurements in the dataset, where an individual patient may have more than one measurement.
No. with diagnosis is the number of rows with a known diagnosis.
No. with IDMS is the number of rows for which creatinine was traceable to an IDMS reference.

Table 4.3 Patient population summary for each of the eight new datasets.

4.2.1 Comparing different methodology for serum creatinine measurement

As discussed in Section 2.4.2, there are two distinguishing characteristics for measured creatinine. First, the method used to measure serum creatinine can either be a Jaffe reaction or an enzymatic method. For the ten datasets, six used a Jaffe method to measure creatinine, *Gothenburg* used an enzymatic method, *Cambridge New* changed from a Jaffe to an enzymatic method, and we did not know the methodology used for the final two centres. Second, creatinine can be IDMS-traceable or not. IDMS-traceable creatinine was implemented in three (*Gothenburg*,

Cambridge new and *Manchester*) out of the ten datasets used in this thesis. For *Gothenburg*, all of the creatinine data were IDMS-traceable, which is in contrast to *Cambridge New* and *Manchester* where IDMS-traceable creatinine was introduced during the data collection period. This combination of different creatinine characteristics in the ten datasets enabled us to compare the differences in serum creatinine for those measurements which were or were not IDMS-traceable, and which were measured using a Jaffe or enzymatic method.

Of the ten datasets, three (*Glasgow*, *Southampton* and *London-Barts*) contained only male patients with germ cell cancer, *Wales* contained only male patients with germ cell cancer or female patients with ovarian cancer, with the remaining centres comprising of a general cancer patient population, some with small subsets of non-cancer patients. Given this, serum creatinine from different centres and measurement methods were compared separately in three groups of patients: male patients with germ cell cancers (n = 1,439), female patients with gynaecological cancers (n = 1,414), and general patients with cancer (n = 7,853). These were chosen so that all centres could be compared while mitigating the bias caused by differences in patient demographics with particular cancer diagnoses in the comparison. Germ cell cancer and gynaecological cancer were also the most common diagnoses for male and female patients respectively. Creatinine values were compared on the natural logarithm scale.

Figure 4.6 compares the creatinine values from the centres which had a general cancer patient population. It is evident that the non-IDMS-traceable values are larger than IDMS-traceable values on average. The mean of log(SCr) across the four centres using non-IDMS-traceable Jaffe was -0.0571 (s.d. 0.0037), which was considerably higher than the mean for the two centres with IDMS-traceable enzymatic (mean -0.225, s.d. 0.0070) and IDMS-traceable Jaffe (mean -0.257, s.d. 0.0097) creatinine values. The differences between all three types of creatinine values were significant, when compared using a t-test. The variability between centres of mean log(SCr) measured using a non-IDMS-traceable method was greater than the IDMS-traceable methods. No significant difference was observed between the two centres with IDMS-traceable enzymatic or IDMS-traceable Jaffe creatinine values, p-value 0.45 and 0.16 respectively. For non-IDMS-traceable, log(SCr) was significantly different when comparing *Cambridge* with the other three centres, p-value < 0.0001 for all.

Examining the centres which changed their methodology during the data collection period, namely *Cambridge* and *Manchester*, enables a direct comparison of the differing methodologies. Figure 4.6 shows the timelines of creatinine values for both of these centres. For *Cambridge*, multiple creatinine values per patient are included. However, the data were thinned so that there

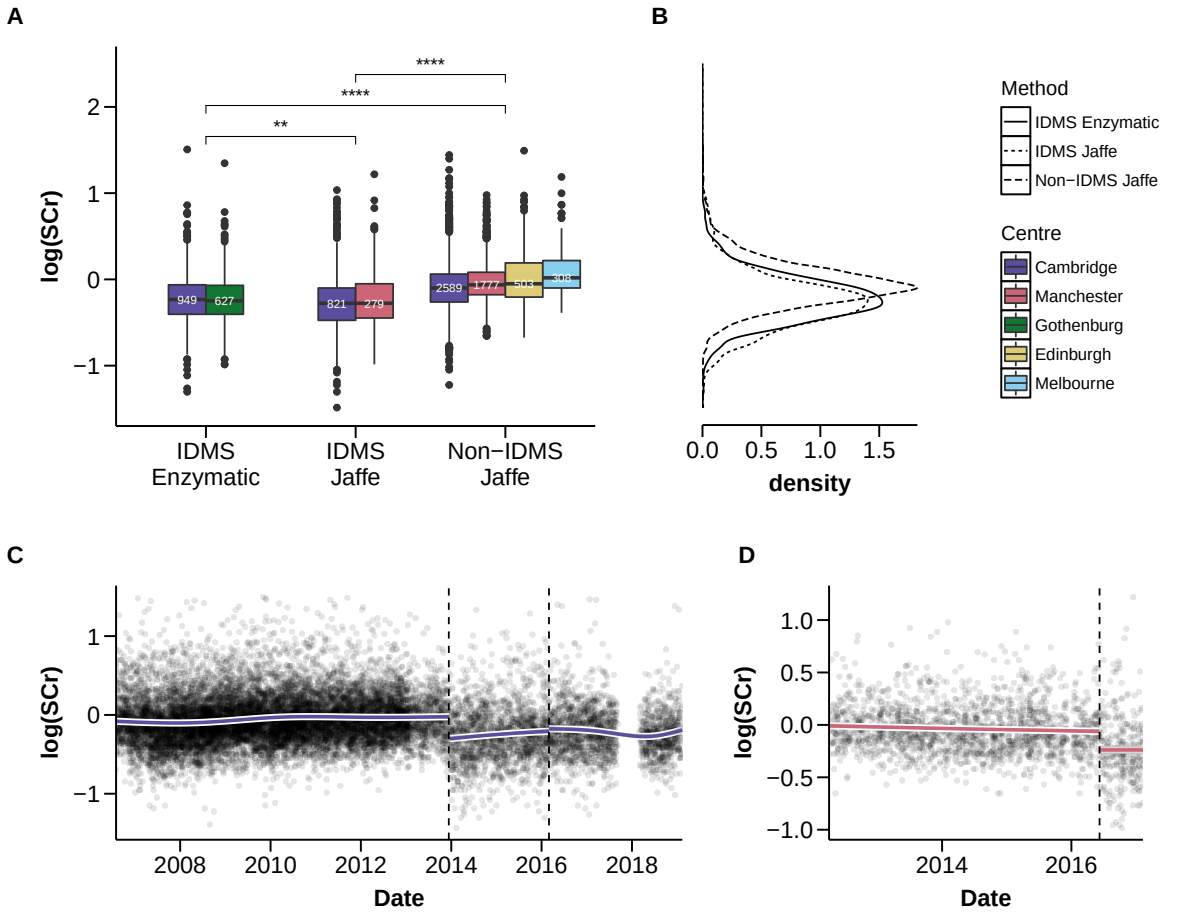


Figure 4.6 Comparison of creatinine methods in patients with cancer. (A) Boxplots for creatinine measurements split by centre and measurement methodology. The numbers in the box correspond to the number of values contained in the given boxplot. The different methodologies were compared with a t-test with ** : $p < 0.01$, **** : $p < 0.0001$. (B) density plot for different creatinine methodologies with the data combined from the different centres in (A). (C) Timeline of creatinine measurements from *Cambridge*. Creatinine is log transformed and data have been thinned so that for any given patient there is a minimum of 30 days between creatinine measurements. The two vertical lines correspond to dates when creatinine measurement methodology changed from non-IDMS-traceable Jaffe to IDMS-traceable Jaffe to IDMS-traceable enzymatic. (D) Timeline of creatinine measurements from *Manchester*. Creatinine is log-transformed and one point corresponds to one patient. The vertical line corresponds to the date when creatinine measurement methodology changed from non-IDMS-traceable Jaffe to IDMS-traceable Jaffe. Smoothed lines are calculated a using generalised additive model with a cubic spline basis.

were at least 30 days between consecutive measurements from any given patient. This was to mitigate the effect of patients having many creatinine measurements in a short period of time when monitoring particularly high creatinine. There is a clear difference in the distribution of log(SCr) when either centres moved from an non-IDMS-traceable Jaffe to an IDMS-traceable

Jaffe method. Moreover, there is a significant difference in the mean log(SCr) value for the non-IDMS-traceable creatinine (p-value < 0.0001) between the two centres but no difference in the mean log(SCr) when both centres were using a IDMS-traceable Jaffe method. Comparing the IDMS-traceable Jaffe with the IDMS-traceable enzymatic creatinine in the *Cambridge* dataset is more difficult. Although overall there is a significant difference in the mean values (p-value 0.002), the mean value appears to fluctuate during the observation period. During the IDMS-traceable Jaffe period there is an increase in the mean value, while during the IDMS-traceable enzymatic period there appears to be an increase in the mean at the end of the observation period. The difference between the two methods appears small either side of the changeover date.

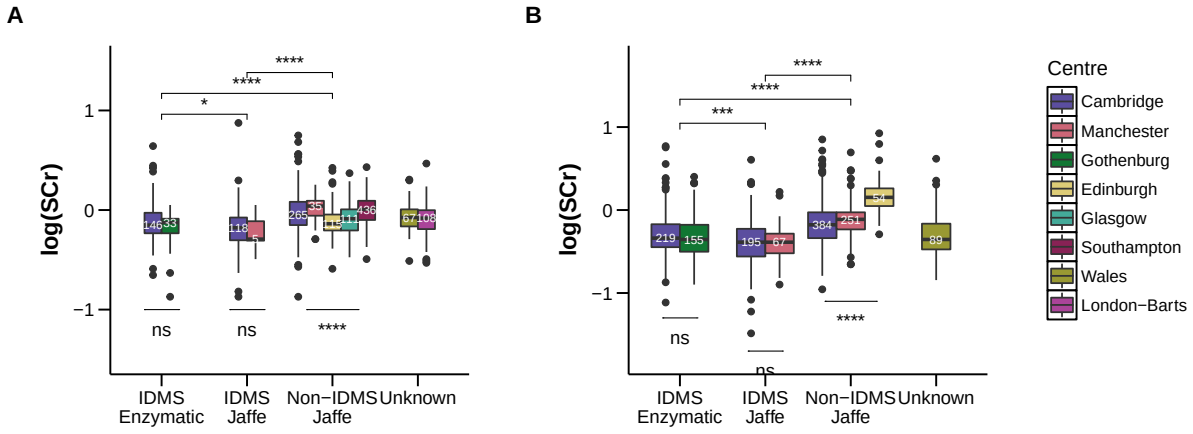


Figure 4.7 Comparison of creatinine methods for patients with gynaecological or germ cell cancers. Boxplots for creatinine measurements split by centre and measurement methodology for male patients with germ cell cancer (A) or female patients with gynaecological cancer (B). The numbers in the box correspond to the number of values contained in the given boxplot. Differences in the mean value between the different methodologies were compared with a t-test (top), differences between centre using a given method were compared using a linear model followed by one-way ANOVA. ns : p < 0.05, * : p < 0.05, *** : p < 0.001, **** : p < 0.0001.

Similar differences between the creatinine methods were observed when the analysis was repeated for male patients with germ cell tumours or female patients with gynaecological cancers. For both, the mean non-IDMS-traceable Jaffe log(SCr) was significantly higher than the IDMS-traceable methods, with inter-centre variability considerably higher when the method was not IDMS-traceable. Additionally, as before, the mean IDMS-traceable Jaffe creatinine was slightly lower than the mean IDMS-traceable enzymatic. A final interesting observation, when comparing non-IDMS-traceable log(SCr) for patients with germ cell cancers in the *Edinburgh* dataset with patients in the *Cambridge* or *Manchester* dataset, the mean is substantially lower.

However, for patients with gynaecological cancer the mean value is substantially higher. The reason for this is not known, however, it may be driven by the observation that the measured GFR values were significantly lower for gynaecological patients in the *Edinburgh* dataset compared with those from *Cambridge* or *Manchester*.

To summarise, from the data there is a clear difference between the mean creatinine value between IDMS-traceable and non-IDMS-traceable values, with IDMS-traceable creatinine being significantly lower on average. There appears to be a small difference between IDMS-traceable Jaffe and IDMS-traceable enzymatic creatinine, with the Jaffe method producing slightly lower values. However, the evidence for this is driven by the *Cambridge* dataset, where the mean log(SCr) appears to fluctuate after the introduction of the IDMS-traceable method.

The differences observed between the IDMS-traceable and the non-IDMS-traceable (or conventional) assays are supported by previous literature. As discussed in Chapter 2 Section 2.4, conventional Jaffe assays were known to overestimate creatinine as a result of picrate reaction with other molecules. Hence, it is not surprising that calibration of these methods reduces the measured creatinine. This reduction was widely acknowledged and discussed, particularly when the National Kidney Disease Education Program first started to recommend that all creatinine should be IDMS-traceable [84, 97, 75]. This difference led to Ortho Clinical Diagnostics producing an equation to convert IDMS-traceable creatinine to "conventional units" [93], which has subsequently been recommended to allow the use of older GFR estimation models developed using conventional creatinine assays.

Previous literature does not fully support the differences observed between the enzymatic and Jaffe methods. A study by the French Society of Clinical Biochemistry compared creatinine determined by use of compensated Jaffe and enzymatic methods from four different assay manufacturers which were all IDMS-traceable [9]. The study found that the Jaffe method was biased to overestimate the true value in the assays from three out of the four manufacturers. Two of these manufacturers were Roche and Siemens, who produced the assays used to measure creatinine in our data (see Table 3.1). The enzymatic assays produced smaller and less biased results than the corresponding Jaffe assay from the same manufacturer. This result is the opposite of what we see from our data, where the Jaffe values were smaller on average compared with enzymatic values. Another study performed in China found similar results, where Jaffe methods were observed to overestimate true creatinine values, whereas enzymatic methods had a tendency to underestimate creatinine but were less biased [94]. Another study which analysed the results from 26 different instruments at five different target serum creatinine

concentrations gave a more complicated picture [55]. Although IDMS-traceable Jaffe methods were biased to overestimate creatinine for low concentrations, for higher concentrations the different methods were more comparable.

Given these findings, in particular the observed differences in serum creatinine between assays which are IDMS-traceable and conventional non-IDMS-traceable assays, it is clear that the published CamGFR model [60] developed in Section 4.1 needs to be adjusted before it is suitable for use with creatinine that is traceable to IDMS reference material. However, given that conventional assays are still routinely used, both in the UK and abroad, it was first important to validate the CamGFR using conventional creatinine that is not IDMS-traceable. The next section compares the CamGFR model with other models using the large multi-centre non-IDMS-traceable creatinine data. Following this, Section 4.3 further develops the CamGFR model to be suitable for serum creatinine traceable to an IDMS-traceable standard.

4.2.2 Validation of CamGFR for non-IDMS-traceable creatinine data

The work in this section is based on our publication, in which the CamGFR model was compared with other previously published models in a large multi-centre dataset [132]. This dataset includes all patients acquired after the publication of CamGFR who had serum creatinine measured using conventional non-IDMS-traceable assays. For some patients, no specific information was known on the methodology used to measure creatinine, however it was suspected that the methods were non-IDMS-traceable and hence these patients were included. In total there were 3,786 patients from seven different centres, of which 3,484 patients had solid cancer, 136 had haematological cancer, and 166 had a non-cancer diagnosis.

As discussed previously, all these patients had creatinine and GFR measured within 30 days of each other. In fact, for 27% of the patients the measurements were performed on the same day and 89% within a week. The median GFR was 85 mL/min ((IQR 61–109 mg/dL) and the median serum creatinine was 0.95 mg/dL (IQR 0.83–1.11 mg/dL). The median age, height, weight and BSA were 59 years, 169 cm, 74 kg and 1.85 m^2 respectively (Table 4.4). These are broadly comparable with the *Cambridge Original* dataset (Table 4.1); the main difference being a higher proportion of male patients (55% compared with 43%) which resulted in a higher median measured GFR, higher median serum creatinine. Of the 3,786 patients, 22 were classified as ethnically black.

	Median (IQR)	Range	Mean (SD)
Age [years]	59 (45-70)	18-91	57 (16)
Weight [kg]	74 (63-86)	33-200	76 (19)
Height [cm]	169 (160-177)	137-204	169 (11)
BSA [m^2]	1.85 (1.68-2.02)	1.17-3.17	1.85 (0.25)
Serum creatinine [mg/dL]	0.95 (0.83-1.11)	0.43-4.45	0.99 (0.28)
GFR [mL/min]	85 (61-109)	9-209	86 (32)

Dataset	Total	Solid tumour	Haematology	Non-cancer	Female	Race - black
Cambridge	404	227	114	63	198	6
Edinburgh	597	472	22	103	245	0
London-Barts	108	108	0	0	0	0
Manchester	1777	1777	0	0	1066	16
Melbourne	308	308	0	0	111	0
Southampton	436	436	0	0	0	0
Wales	156	156	0	0	89	0
Total	3786	3484	136	166	1709	22

GFR was measured using either 99mTc-DTPA (Edinburgh and Melbourne) or 51Cr-EDTA (all others).
GFR - Glomerular filtration rate, BSA - Body surface area (calculated using DuBois-DuBois [33]), IQR - interquartile range
Modified from Table 1 in [132]

Table 4.4 Patient characteristics for CamGFR validation data. Continuous data are summarised for all centres together (top) and categorical data are split by dataset (bottom).

Figure 4.8 compares the distributions of the continuous variables for the patients from the seven different centres. Patients from *Southampton* and *London-Barts* have a higher GFR, height and weight and a lower age than patients from other centres. This is due to them comprising male patients with germ cell tumours. Patients from *Manchester* and *Melbourne* were typically older compared with other centres. Patients from *Manchester* were shorter, weighed less and had lower measured GFR too. This is may be driven by the higher proportion of female patients (60% compared with 44% for the other mixed centres, Table 4.4). Inter-centre comparisons of log(SCr) were discussed in the previous section.

The same published models compared in Section 4.1 were again compared. As discussed in Chapter 3, Section 3.2.3 the statistics used to compare models were changed slightly between the two publications. Here, models were compared using RMSE, residual IQR and median residual, which assessed the models' accuracy, precision and bias respectively. A final statistic, denoted P20, which is the proportion of patients with an absolute percentage error more than 20%, was used as a measure of the clinical robust measurement of accuracy. Bootstrap resampling methods were used to calculate approximate 95% confidence intervals [25] and to calculate p-values [36] for comparing statistics from different models, further detailed in

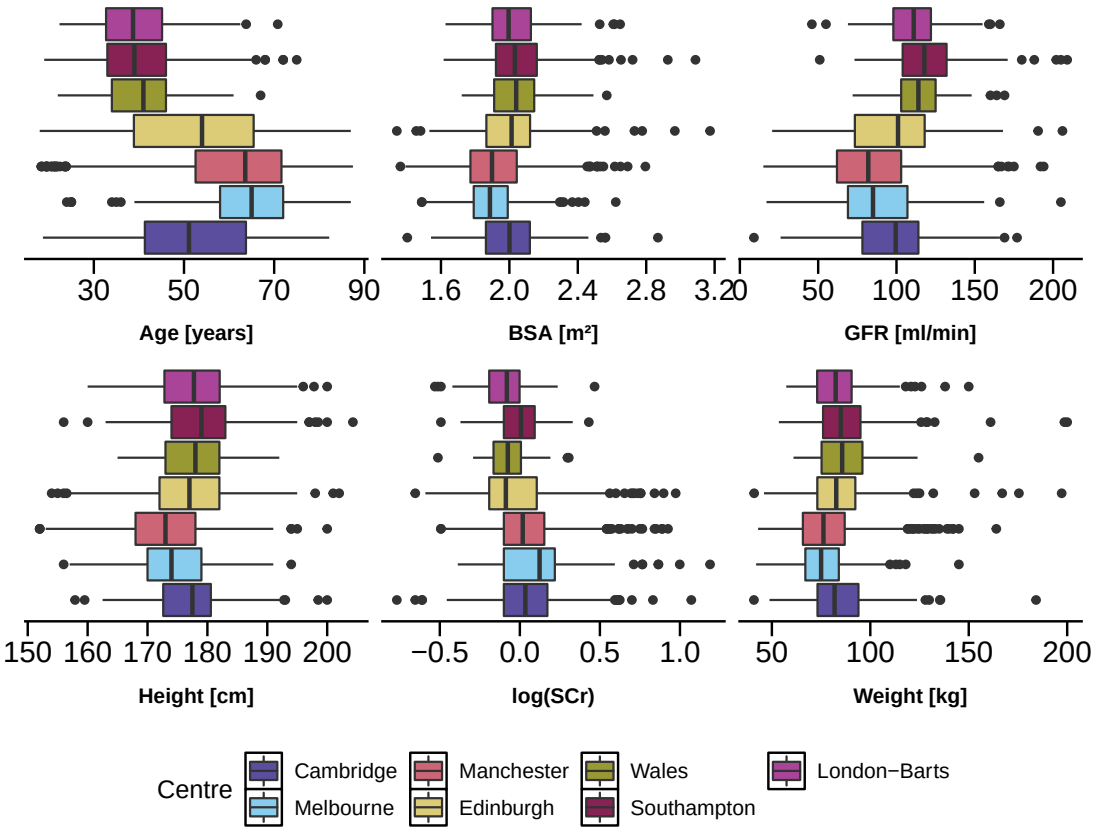


Figure 4.8 Boxplot of the continuous variables (age, BSA, GFR, height, weight, log(SCr)) for each centre used to validate the CamGFR model.

Section 3.2.3. Models were compared on the whole patient population as well as in relevant sub-populations, such as for each centre separately.

Overall, CamGFR was significantly more accurate in estimating GFR than all other models, both by RMSE or P20, followed by the CKD-EPI model (Figure 4.9). The RMSE for the CamGFR model was 17.3 mL/min (CI 16.7 to 17.8 mL/min) compared with 18.1 mL/min (CI 17.6 to 18.7 mL/min) for the CKD-EPI model (p-value = 0.03). For P20, the value for CamGFR was 0.295 (CI 0.281 to 0.310), or in other words 29.5% of patients' estimated GFR differed by more than 20% of their true GFR. This compares with 0.318 (CI 0.304 to 0.333 for CKD-EPI), which was significantly higher than the CamGFR (p-value = 0.03).

The residual median was 1.30 mL/min (CI 0.74 to 1.8 mL/min) for CamGFR and 0.60 mL/min (CI 0.08 to 1.11 mL/min) for CKD-EPI, indicating a small but significant bias to overall under-

estimation of GFR in both models. For the MDRD and Cockcroft-Gault, the residual median was not significantly different from zero. For the other four models (Wright, Jelliffe, Martin and Mayo) large biases were observed, with the residual median equal to -4.86, 9.37, -7.89 and -16.77 respectively. Since the RMSE of Jelliffe, Martin and Mayo is larger than $22\,\mathrm{mL/min}$, these models were not compared further.

When comparing models in individual centres, the CamGFR was most accurate in six of the seven centres. The exception was *Melbourne*, where the Wright model was the most accurate. The accuracy of all models fluctuated between models, with the RMSE for the CamGFR ranging from $15.7\,\mathrm{mL/min}$ for patients from *Cambridge* to $23.5\,\mathrm{mL/min}$ for patients from *Cambridge*. This accuracy was partially driven by centre-dependent bias for serum creatinine.

All models were generally biased to overestimate GFR for patients from *Edinburgh*, *Wales* and *London-Barts*, and to underestimate GFR for patients from *Southampton* and *Melbourne*. This is evidenced by the sign of the median and the confidence interval not overlapping zero, with a positive median residual indicating underestimation. The CamGFR, CKD-EPI and MDRD had no significant bias for patients from *Cambridge New*. This centre-specific bias can be correlated to a degree with previous observations of centre-specific differences in serum creatinine, see Figures 4.6 and 4.7. More specifically, the creatinine for patients from *Southampton* and *Melbourne* was higher compared with centres with a similar patient population. In particular compared with creatinine from *Cambridge* which was used to develop the CamGFR and showed no bias. Thus, given that a higher creatinine would result in a lower estimated GFR, these centres would have lower estimated GFR values. The observed underestimation in the *Edinburgh*, *Wales* and *London-Barts* is harder to explain given the data. Given the previous observation, we might expect these centres to have lower creatinine compared with the unbiased centres. However, this is not the case for patients from *Edinburgh* overall, patients from *London-Barts* or germ cell patients from *Wales*. It is the case for germ cell patients from *Edinburgh* and gynaecological patients from *Wales*.

In further subgroup analysis, we examined the performance of models for patients with solid cancer, with haematological cancer, and without cancer. CamGFR was the most accurate for each of these groups, RMSE was 17.3, 18.7 and $14.1\,\mathrm{mL/min}$ for CamGFR compared with 18.1, 21.3 and $15.8\,\mathrm{mL/min}$ respectively for the CKD-EPI model. The accuracy of CamGFR in patients without cancer is of particular interest as it was developed using a cancer patient population yet it accurately predicts GFR in non-cancer patients.

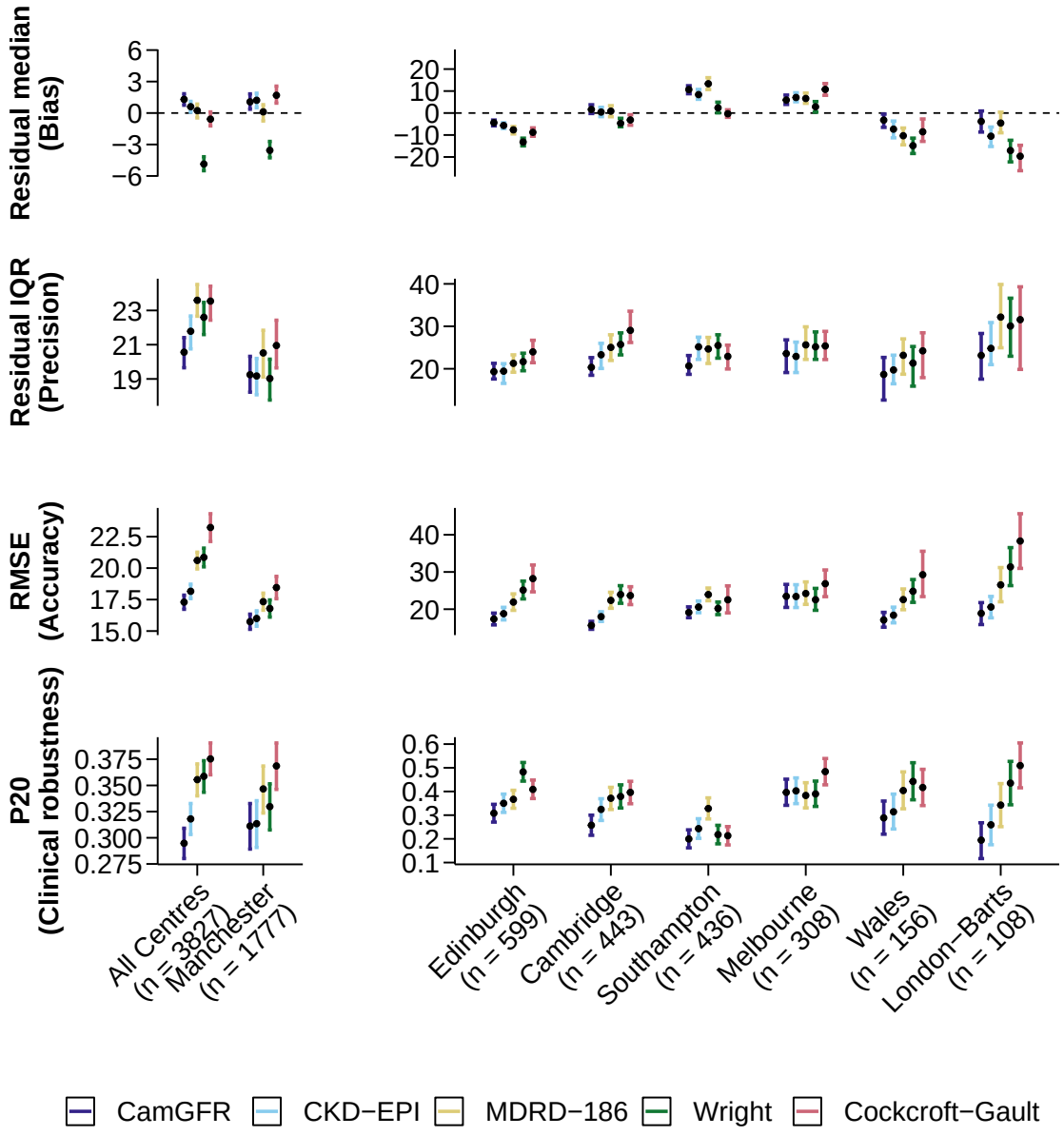


Figure 4.9 Summary statistics comparing CamGFR with commonly used and well performing models. Results for the five best-performing models (CamGFR, CKD-EPI, Wright, MDRD-186 and Cockcroft-Gault) for the 3,268 patients from the non-IDMS-traceable creatinine validation dataset are displayed. Statistics are shown for all centres and for the individual centres separately. The residual (measured GFR - estimated GFR) median (first row), residual interquartile range (second row), root-mean-squared error (third row) and the proportion of patients who have an absolute percentage error more than 20% (P20) (fourth row) are displayed. All error bars are 95% confidence intervals calculated using bootstrap resampling with 2,000 repetitions and a normal distribution approximation. Figure is adapted from Figure 1 in [132]

Differences in accuracy were observed between subgroups split by estimated GFR, age, creatinine, or BSA. However, these are likely a result of calculating residuals using untransformed

GFR and not the transformed scale on which it was modelled. For the CamGFR a square-root transformation was used, for the other models typically a log transformation is used. This results in trends for larger absolute residuals for larger GFR values. Figures for all subgroup analyses are available in [132].

One of the strengths of the CamGFR model is the possibility to calculate the prediction intervals for the estimated GFR value, which can be set at any predefined level. The statistical consensus and therefore the clinical consensus is to define a 95% confidence interval. To assess the reliability of the prediction intervals across the different centres the 95% interval was calculated for all the patients in the validation dataset. Out of the 3,786 patients, 3,510 (92.7%) had measured GFR values within their prediction interval, which is slightly lower than the expected 95% value. Figure 4.10 depicts the prediction intervals for the patients from each centre separately, and provides the percentage of measured GFR values within the 95% prediction interval for each cancer centre. The previous bias (observed in Figure 4.9) is once again observable in this plot. In particular, patients from *Southampton* and *Melbourne* tend to have higher measured values than estimated values. However, 95.4% of patients from *Southampton* have their measured GFR within the 95% prediction interval whereas for *Melbourne* it is only 85.7%.

The work up to this point has demonstrated that using non-IDMS-traceable creatinine in the CamGFR model can reduce the RSME and therefore improve the accuracy of estimated GFR compared with previously published models. This was observed in the multi-centre analysis but significant biases for individual centres were also observed. These biases may in part be explained by the different methods used to measure creatinine, discussed in Section 4.2.1. We further observed that the average creatinine value was reduced when it was IDMS-traceable compared with when it was not. This led to the conclusion that the CamGFR would not be suitable if IDMS-traceable creatinine was used as an input. In the next chapter, we will expand the CamGFR to be suitable for IDMS-traceable creatinine and in doing so explore whether a more accurate model can be developed using the additional data collected.

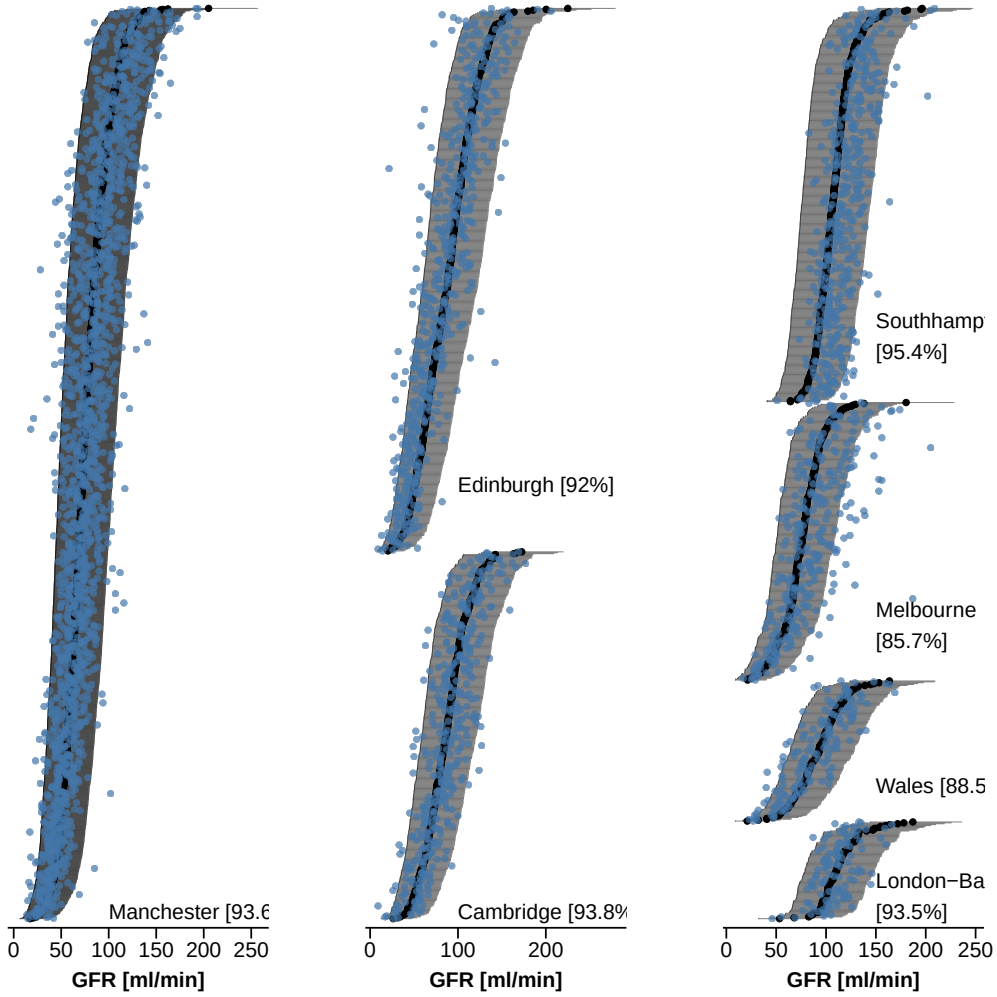


Figure 4.10 Confidence interval for estimated GFR of each patient used for the CamGFR model validation. The measured GFR (blue points) and the estimated GFR (black points) for each patient are displayed. Each horizontal line represents a 95% prediction confidence interval for the patient, with patients grouped by their cancer centre and ordered in accession by their estimated GFR. For each centre the percentage of patients whose measured GFR is within the 95% confidence interval is given.

4.3 Refitted CamGFR: Expanding CamGFR for IDMS-traceable creatinine data

In Section 4.2.1, we observed a significant difference in the average value of serum creatinine that was or was not traceable isotope dilution mass spectrometry (IDMS) standardised. Given this difference, if IDMS-traceable creatinine were used with the CamGFR model, the estimated

GFR would be biased compared with measured GFR. Hence, further refinement was required for the CamGFR model to be suitable for IDMS-traceable creatinine. This section describes the redevelopment of the CamGFR model to be suitable for both IDMS-traceable and non-IDMS-traceable creatinine. Furthermore, the currently available datasets contains more patients from multiple centres, so that redeveloping the model should produce more accurate and less centre-dependent coefficients.

Several approaches were considered to redevelop the CamGFR model. First, IDMS-traceable creatinine values could be normalised to have the same distribution as the non-IDMS-traceable creatinine. These normalised creatinine values could then used as the input for the CamGFR. This adjustment approach was suggested by Ortho Clinical Diagnostics [93], who produced the equation:

$$\text{non-IDMS-traceable creatinine (mg/dL)} = \text{IDMS-traceable creatinine (mg/dL)} \times 1.065 + 0.067 \tag{4.4}$$

to convert between IDMS-traceable and non-IDMS-traceable creatinine when using their analysers. This equation or a similar equation could be used to transform the new IDMS-traceable creatinine to have the same distribution as the non-IDMS-traceable creatinine for our data. Then these data could be used with the CamGFR model developed in Section 4.1, thus mitigating the need to adjust the model. However, this approach has some issues: it does not utilise the additional data we now have, we cannot be sure that the adjustment equation we use will be accurate for unseen data from other cancer centres, and it does not resolve the possible centre-dependence of the CamGFR model. Given these limitations, this solution was not pursued.

Our first approach to adjusting the CamGFR was to include an additional variable "creatinine-type", which indicated whether the creatinine was IDMS-traceable or non-IDMS-traceable, along with interaction terms between this variable and the cubic creatinine variable. The resulting model would have the same variables as the original CamGFR, but different coefficients. Additionally, the coefficients for the creatinine terms would differ based on creatinine-type. This would allow some consistency with the original CamGFR. However, it relied on the assumption that the same variables are informative in predicting GFR when using multi-centre data.

To address the above concerns, we also redeveloped the model from the beginning using the multi-centre dataset. For this and the previous approach, data were used from the centres which

had at least some IDMS-traceable creatinine data (*Cambridge*, *Manchester* and *Gothenburg*); this included the original *Cambridge* data used to develop CamGFR. These three centres contained 8,243 GFR measurements from 7,239 patients. In addition to the exclusion criteria in Section 3.1.3 we further excluded samples if they were a repeat GFR for the same patients within a year of the last measurement (n = 448). This was to mitigate for any adverse effect of treatment and loss of independence of the samples. After these samples were removed, there were 7,795 GFR measurements from 7,238 patients, of which 3,182 had IDMS-traceable creatinine. The data from these centres were subsequently randomly split into development and validation datasets at the ratio 4:1.

4.3.1 Refitting the CamGFR model

As discussed above, our first approach was to adjust the CamGFR model with an interaction variable between the creatinine variables and creatinine-type. Hence, the regression model has the following form:

$$\begin{aligned}\sqrt{\text{GFR}} = {} & \beta_0 + \beta_1 \text{Age} + \beta_2 \text{BSA} + \beta_3 \text{Sex}_\text{M} + \beta_4 \text{Scr}_\text{IDMS} \\ & + \beta_5 \log(\text{Scr}) \text{Scr}_\text{IDMS} + \beta_6 \log(\text{Scr})^2 \text{Scr}_\text{IDMS} + \beta_7 \log(\text{Scr})^3 \text{Scr}_\text{IDMS} \\ & + \beta_8 \log(\text{Scr}) \text{Scr}_\text{non-IDMS} + \beta_9 \log(\text{Scr})^2 \text{Scr}_\text{non-IDMS} \\ & + \beta_{10} \log(\text{Scr})^3 \text{Scr}_\text{non-IDMS} + \beta_{11} \text{AgeBSA} + \beta_{12} \text{AgeSex}_{\text{M}+\varepsilon}\end{aligned} \tag{4.5}$$

where

- β_i for $i = 0, \ldots, 12$ are the coefficients estimated via least squares
- BSA is the patient's body surface area
- Sex_M is a dummy variable equalling 1 if the patient's sex is male and 0 if female
- Scr_IDMS is a dummy variable equalling 1 if the creatinine measurement is IDMS-traceable and 0 if it is not
- $\text{Scr}_\text{Non-IDMS}$ is a dummy variable equalling 1 if the creatinine measurement is non-IDMS-traceable and 0 if it is not
- Scr is the patient's serum creatinine measurement
- ε are mean zero and constant variance residual

This model will be referred to as the refitted CamGFR model. The coefficients estimated from fitting the model using the development dataset are given in Table 4.5. Some coefficients ($\beta_0, \beta_1, \beta_2, \beta_3, \beta_8, \beta_9, \beta_{10}, \beta_{11}, \beta_{12}$) are directly comparable with the coefficients from the original CamGFR model, where the variable $\text{Scr}_{\text{Non-IDMS}}$ was equal to 1. Hence, Table 4.5 additionally displays the original coefficients, their standard error, and a p-value resulting from a Z-test for the difference between the two sets of coefficients. Of the nine comparable coefficients, only $\log(\text{Scr})\text{Scr}_{\text{IDMS}}$ had a significantly different coefficient in the refitted CamGFR model. This suggests that both models are reliable and have not over-fitted the data. All coefficients in the refitted CamGFR model were significantly different from zero, indicating that all coefficients are necessary.

Variable	Coefficient	New	New SE	Original	Original SE	p-value*
(Intercept)	β_0	1.662	0.374	1.814	0.626	0.781
Age	β_1	0.018	0.006	0.019	0.010	0.793
BSA	β_2	4.772	0.213	4.733	0.355	0.794
Sex_{M}	β_3	0.302	0.102	0.020	0.174	0.301
Scr_{IDMS}	β_4	-0.508	0.029			
$\log(\text{Scr})\ \text{Scr}_{\text{IDMS}}$	β_5	-3.499	0.084			
$\log(\text{Scr})^2\ \text{Scr}_{\text{IDMS}}$	β_6	-0.738	0.092			
$\log(\text{Scr})^3\ \text{Scr}_{\text{IDMS}}$	β_7	0.699	0.112			
$\log(\text{Scr})\ \text{Scr}_{\text{Non-IDMS}}$	β_8	-4.049	0.077	-3.716	0.086	0.012
$\log(\text{Scr})^2\ \text{Scr}_{\text{Non-IDMS}}$	β_9	-1.162	0.115	-0.914	0.117	0.253
$\log(\text{Scr})^3\ \text{Scr}_{\text{Non-IDMS}}$	β_{10}	1.532	0.160	1.063	0.132	0.062
Age BSA	β_{11}	-0.028	0.004	-0.030	0.006	0.765
Age Sex_{M}	β_{12}	0.006	0.002	0.012	0.003	0.125

*p-value calculated for the null hypothesis that the coefficients for the two different regression fits are equal [21]

Table 4.5 Comparison of coefficients form the original CamGFR and refitted CamGFR models.

Alternatively, the fit of a regression model can be assessed with the help of summary plots, as seen in previous sections. Figure 4.11 plots the residuals of the model (on the square root scale) against the three main continuous variables along with the fitted values. Any trends in these plots indicate that the model may not be fully explaining the relationship in the data. In Figure 4.11, no major patterns are seen. However, there does appear to be a small bias to overestimate GFR for patients who are young, indicated by the deviation of the fitted line from a residual of 0. The other observed deviations appear to be driven by relatively few samples and hence are less interesting.

We used the validation dataset to compare estimated GFR from the original and refitted CamGFR models, Table 4.6. Patients with IDMS-traceable or non-IDMS-traceable creatinine

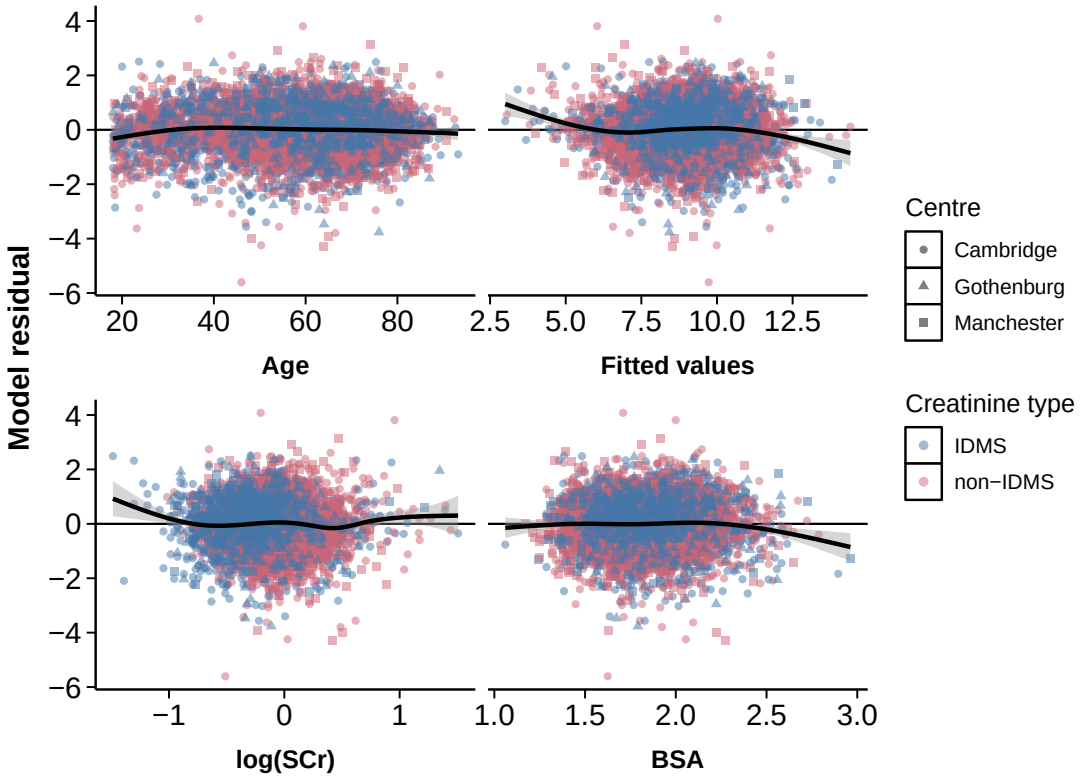


Figure 4.11 Model residuals against fitted values and three explanatory variables (age, log(SCr) and BSA) for the refitted CamGFR model. Residuals and fitted values are shown on the square-root scale. For each patient the colour indicated creatinine-type (IDMS or non-IDMS) and the shape indicated the cancer centre. All data smoothing lines are loess curves with the span parameter set to 0.9.

were analysed separately. For non-IDMS-traceable creatinine the refitted and original CamGFR were broadly comparable in terms of accuracy, precision and bias. The refitted CamGFR produced slightly higher values for estimated GFR and was more precise and accurate for data from both centres. However, it was more biased for patients from *Cambridge*, with a tendency to overestimate GFR. This is likely due to it being fitted using data from multiple centres and the inter-centre variability of non-IDMS-traceable creatinine. Overall, the refitted CamGFR was slightly more accurate although the difference is insignificant. As expected, for IDMS-traceable creatinine the original model is biased to overestimate GFR. This bias was prevalent for patients from all centres but was particularly high for patients from *Cambridge* where on average a patient GFR would be overestimated by 9.07 mL/min. Overall, the refitted CamGFR was not biased; however, for patients from *Gothenburg* there was a bias to underestimate GFR on average. This is an unexpected result particularly as there was no difference in creatinine

model	Centre	n	median value	RMSE	median residual	IQR
IDMS creatinine						
Original	All	618	91.05	17.02 (15.86, 18.20)	-8.2 (-9.8, -6.96)	18.23 (16.89, 20.30)
Refitted	All	618	81.79	14.97 (13.84, 16.13)	0.73 (-0.68, 2.20)	17.79 (15.78, 19.07)
Original	*Cambridge*	426	92.95	17.58 (16.17, 19.05)	-9.07 (-10.22, -7.75)	17.32 (15.43, 19.41)
Refitted	*Cambridge*	426	82.94	14.90 (13.44, 16.40)	-0.37 (-1.85, 1.45)	18.26 (16.60, 20.70)
Original	*Gothenburg*	130	87.23	15.78 (13.23, 18.27)	-3.8 (-7.72, -0.17)	18.05 (14.81, 21.32)
Refitted	*Gothenburg*	130	78.79	15.36 (13.13, 17.55)	5.21 (2.40, 8.68)	17.77 (14.37, 20.13)
Original	*Manchester*	62	86.54	15.54 (12.11, 19.22)	-5.84 (-9.55, -2.23)	15.47 (11.02, 20.33)
Refitted	*Manchester*	62	77.74	14.61 (11.73, 17.70)	1.92 (-2.45, 5.27)	17.00 (13.40, 22.70)
Non-IDMS creatinine						
Original	All	921	78.63	15.83 (14.95, 16.72)	-0.09 (-1.25, 0.92)	19.10 (17.63, 20.95)
Refitted	All	921	79.01	15.74 (14.86, 16.63)	-0.43 (-1.48, 0.91)	18.55 (17.17, 19.93)
Original	*Cambridge*	577	80.45	15.28 (14.16, 16.38)	-0.38 (-1.84, 1.14)	18.95 (16.99, 21.14)
Refitted	*Cambridge*	577	81.36	15.41 (14.29, 16.52)	-1.50 (-3.08, -0.18)	18.47 (16.37, 20.43)
Original	*Manchester*	344	76.50	16.72 (15.25, 18.23)	0.48 (-1.12, 2.05)	19.36 (16.49, 22.17)
Refitted	*Manchester*	344	75.77	16.28 (14.84, 17.76)	0.33 (-1.84, 1.95)	18.66 (15.99, 20.94)

Table 4.6 Comparison of the refitted and original CamGFR models for the validation dataset. Patients with IDMS-traceable and non-IDMS-traceable creatinine are compared separately.

between *Cambridge* and *Gothenburg*, Figure 4.6. The refitted CamGFR also had the smallest RMSE and residual IQR.

To conclude, the refitted CamGFR model fits the data well. It is comparable with the original CamGFR model for non-IDMS-traceable data and far more accurate for IDMS-traceable creatinine. The next section will investigate whether further improvements can be achieved by adjusting the model.

4.3.2 Adjusting the CamGFR model

This section will present the analysis for redeveloping a model to estimate GFR. The analysis will follow a similar structure to the original development of the CamGFR model discussed in Section 4.1. Specifically, we first investigate refitting a linear model, with a transformation applied to the response variable, GFR. Following this we consider whether there is any additional benefit in modeling the GFR-creatinine and GFR-age relationships with a piecewise linear fit. Generalised linear models were not considered due to complexities in producing confidence intervals for the predicted values.

Assessing relationships in the data

As in Section 4.1, we calculated the Box-Cox transformation. Using the initial model, given in Equation 4.1, with the addition of the previously discussed interaction term between creatinine and creatinine-type, the power transformation calculated was 0.578 (95% CI 0.538–0.619). Although the confidence interval does not contain 0.5 (or the square root transformation), it is reasonably close. Hence, as with the original CamGFR, this transformation will be used.

Analysis of the relationship between square-root-transformed GFR and the explanatory variables gave similar observations to those in Figure 4.1. As before, creatinine was log-transformed to improve linearity and reduce the skewness. A deviation from linearity in the relationship between age and sqrt(GFR), where the slope of the relationship decreases for young patients, was more prominent in this dataset compared with Figure 4.1. No difference was observed between the IDMS-traceable and non-IDMS-traceable patient groups for the non-creatinine variables. For a subset of the data, ethnicity was annotated with the majority of patients annotated as being ethnically white (2,093 out of 2,205 for the development dataset). There was an observable difference in the GFR-creatinine relationship between the 21 ethnically black and the ethnically white patients, where ethnically black patients had a higher GFR for the same creatinine value. However, due to the small number of ethnically black patients, this variable was not considered for the model due to difficulty in estimating accurate coefficients.

Stepwise model selection

Stepwise selection, using the three different criteria of five-fold cross validation, leave-one-out cross validation, and Bayesian information criterion, were again used. The set of models searched contained all the combinations of the variables for age, sex, height, weight, BSA, log(SCr), creatinine-type along with the pairwise interaction between these variables.

A few small changes were made to the procedure for the five-fold cross validation criteria. The model selection algorithm was altered to start from the null model once in every 100 repetitions. For the remaining repetitions the algorithm starts from the previously selected model. This reduced the computation time to run the algorithm. The number of repetitions was also increased from 2,000 to 5,000.

Unlike for the original CamGFR model the three criteria selected slightly different models. Figure 4.12 summarises the variables included for each selection methods and compares fitted coefficients with the coefficients for the refitted CamGFR model. All three models included several new variables that were not included in the refitted CamGFR model. These included weight and height variables along with the BSA-sex and BSA-height interaction terms, with the leave-out-n model also including an age-height interaction term. The BIC model was the only one to retain the age-sex interaction term included in the CamGFR model. All the coefficients for the additional interaction term were significantly different from zero. Only one of these models was taken forward for further consideration, and as the leave-one-out model was the simplest (contained the fewest terms) of the three, it was chosen.

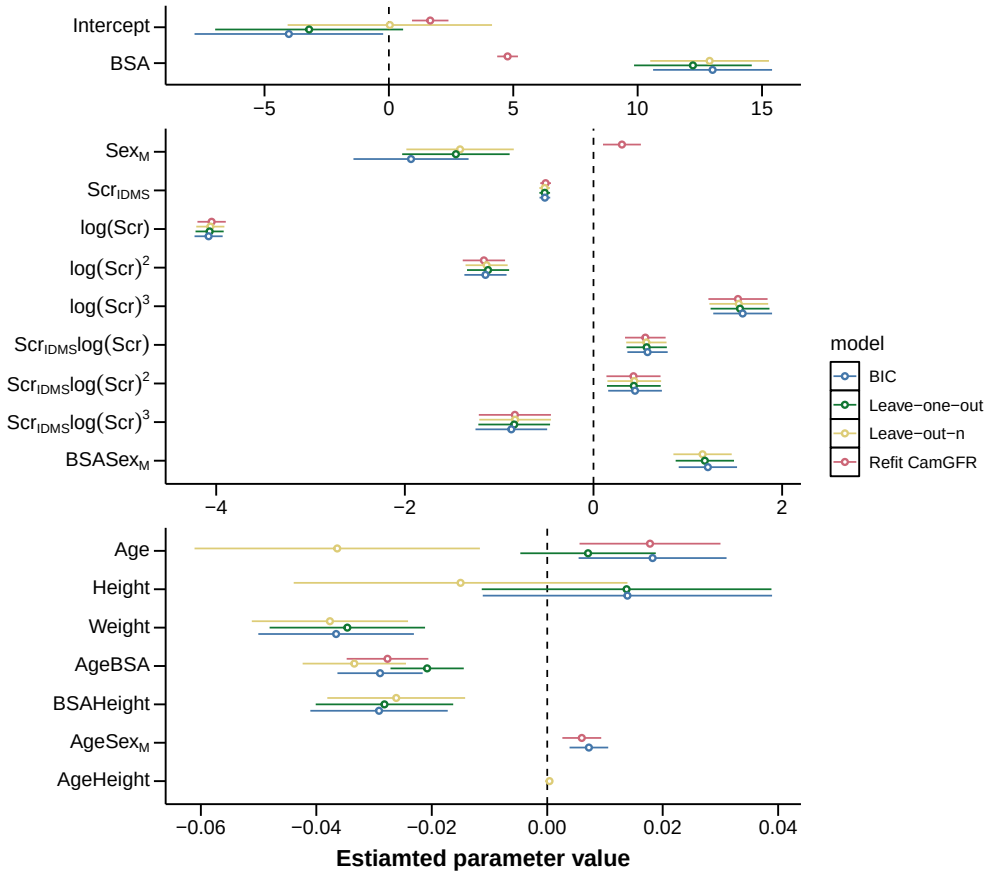


Figure 4.12 Comparison of the fitted coefficients for the BIC, leave-one-out, leave-out-n and refitted CamGFR. For each term included in any of the models the point estimate along with a 95% confidence intervals is shown. Coefficients are split into three groups based on their size and standard error to allow comparison. If the error bar crosses 0 then this indicates that the coefficient is not significantly different from 0.

Segmented regression model

In addition to the above analysis, we examined whether a continuous piecewise linear relationship might be more appropriate for the age and creatinine variables, also known as segmented regression model. This was driven by previous observations, where the relationship between age and GFR appear to be piecewise linear with a break point around 40 years. A piecewise linear curve may also fit the GFR and creatinine relationship with two break points. In addition, the CKD-EPI models the creatinine-GFR relationship with a piecewise linear fit.

To assess whether using a piecewise linear relationship would improve the model, a new model was produced which included a piecewise linear relationship for age and creatinine. Models that only added a piecewise linear relationship for age or creatinine were also considered but were less accurate and thus will not be discussed here. The model was fitted using the **segmented** R package [83], which, given a linear model and the explanatory variables for which a piecewise relationship is desired, estimates the break points and slope of each piece by minimising the log-likelihood. For this, initial values need to be supplied for the break points; these were chosen to be 40 years for age and -0.1 and 0.5 for log(SCr). Figure 4.13 displays the fitted effect curves for log(SCr) and age and compares it with the corresponding effect curves for the refitted CamGFR model. For the age, the estimated break point was 36.9 years, while for log(SCr) the break points were estimated to be -0.137 and 0.504 for IDMS-traceable creatinine and -0.164 and 0.542 for non-IDMS-traceable creatinine. Interestingly, there was not a large shift between the piecewise linear curves for IDMS-traceable and non-IDMS-traceable creatinine, particularly compared with that observed for the cubic fit in the refitted CamGFR model.

When examining diagnostic plots for the newly fitted model, the small overestimate of GFR in younger patients, seen in Figure 4.11, was removed. The plot of residual against log(SCr) also showed some improvements. With these three models, along with the BIC and leave-out-n and refitted CamGFR model, we move forward and compare them with the published model using our validation dataset.

Model comparison

The models fitted in the previous sections, namely the refitted CamGFR, the leave-one-out and the piecewise linear model, were then compared with each other using the validation

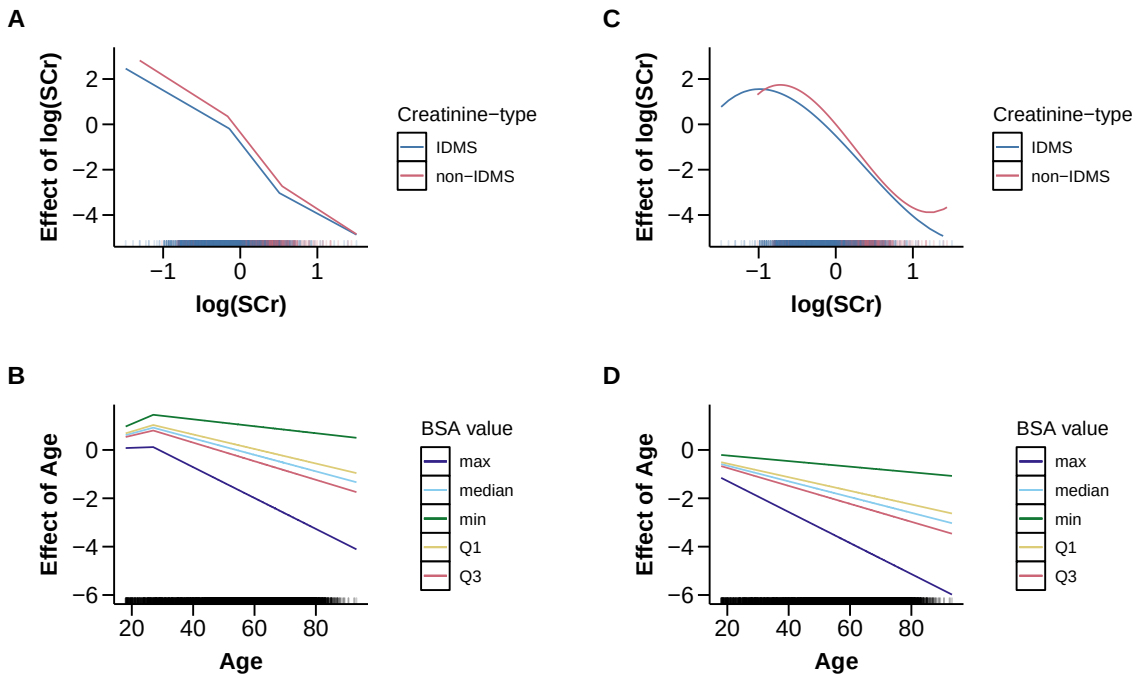


Figure 4.13 Fitted effect curves for log(SCr) and age for the refitted CamGFR (C and D) and the model fitted with piecewise linear relationships for both log(SCr) and age (A and B). The curve for log(SCr) is plotted for IDMS-traceable and non-IDMS-traceable separately. As age has an interaction term with BSA, the curve for age is plotted for the minimum, 1st quartile, median, 3rd quartile and maximum values of BSA in the dataset. The values of the variables for the patients used in the model development dataset can be seen on the x-axis.

dataset. Both IDMS-traceable and non-IDMS-traceable creatinine data were used for this comparison. Table 4.7 provides the summary statistics for the refitted CamGFR, leave-one-out, and piecewise linear models for patients in the validation dataset. All models have similar results for the four statistics used to compare the models. For RMSE, the piecewise linear model was slightly more accurate than the leave-one-out model which itself was slightly more accurate than the refitted CamGFR. However, these differences and the differences between any statistics were not significant, as calculated using the bootstrap procedure described in Section 3.2.3. All three models were unbiased. These small improvements of the piecewise linear and leave-one-out models were not deemed sufficient to warrant changing the form of the CamGFR model, thus we chose the refitted CamGFR as our final model and subsequently compared this model with previously published models.

The previously published models that we compared with the refitted CamGFR were the CKD-EPI [77] and MDRD-186 [76] which were compared during the original development of

model	median value	RMSE	median residual	IQR	P20
Piecewise linear model	80.67	15.28 (14.58, 15.98)	-0.16 (-1.09, 0.76)	18.14 (16.92, 19.21)	0.25 (0.23, 0.27)
Leave-one-out model	80.29	15.31 (14.62, 16.02)	-0.05 (-0.88, 0.81)	18.31 (17.44, 19.19)	0.25 (0.23, 0.27)
Refitted CamGFR	80.03	15.44 (14.74, 16.15)	-0.07 (-0.89, 0.73)	18.54 (17.61, 19.60)	0.25 (0.23, 0.27)

Table 4.7 Refitted CamGFR, leave-one-out and piecewise linear models comparison on the IDMS-traceable and non-IDMS-traceable combined data in the validation dataset.

CamGFR (Section 4.1), along with the Lund-Malmo [7] and FAS [102] models. For this comparison, patients with IDMS-traceable or non-IDMS-traceable were compared separately as all four previously published models were developed using IDMS-traceable creatinine and hence may be biased for non-IDMS-traceable creatinine. Figure 4.14 displays the four statistics used to compare models for each of the models.

The refitted CamGFR was the most accurate and precise for both IDMS-traceable and non-IDMS-traceable creatinine. For IDMS-traceable creatinine the refitted CamGFR had significantly better RMSE than all models (p-value < 0.005) apart from the Lund-Malmo (p-value 0.26). For non-IDMS-traceable data, it was significantly better than all models (p-value $<$ 0.005) apart from the CKD-EPI (p-value 0.06).

The refitted CamGFR was the only model that was unbiased for both IDMS-traceable and non-IDMS-traceable creatinine. This was not surprising given that it is the only model to contain a factor differentiating the two creatinine types. What is more surprising, but perhaps not unexpected given that it was unbiased in validation of the original CamGFR, was the high bias and poor accuracy of the CKD-EPI model on IDMS-traceable creatinine. The CKD-EPI is widely recommended as the best model to estimate GFR when creatinine is IDMS-traceable [11, 86]. However, these results suggest that this could lead to an overall bias to overestimate GFR by 11.3 mL/min when using the CKD-EPI model. There is a clear difference in accuracy and bias between the refitted CamGFR and Lund-Malmo compared with the CKD-EPI and the other two models. There are numerous possible reasons for this discrepancy, the most likely of which is that the IDMS-traceable data used to develop the CKD-EPI (and MDRD) are different (or biased) from the IDMS-traceable creatinine data used in this study.

Analysing CamGFR as date difference increases

Throughout this chapter we have only included patients if they have a serum creatinine measurement within 30 days of their nmGFR measurement. This cut-off was originally chosen

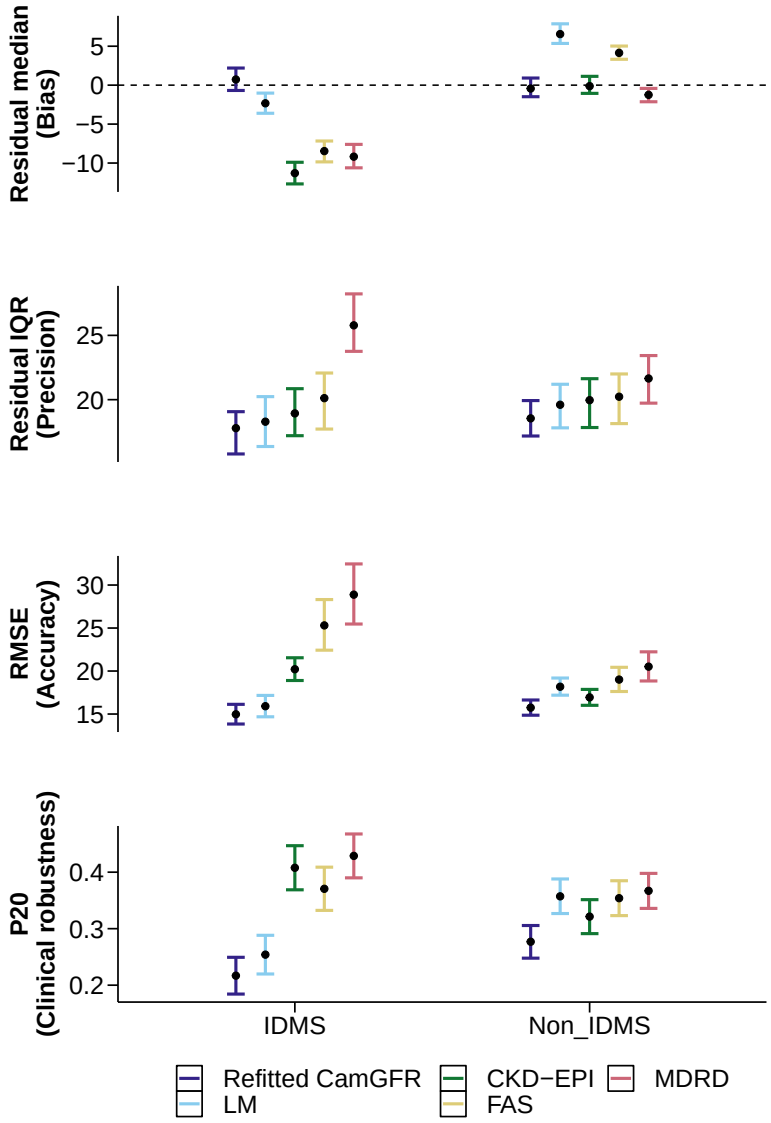


Figure 4.14 Summary statistics comparing the Refitted CamGFR, Lund-Malmo, CKD-EPI, FAS and MDRD models. Statistics calculated for the 618 patients with IDMS-traceable creatinine and 921 patients with non-IDMS-traceable creatinine separately. The residual (measured GFR - estimated GFR) median (first row), residual interquartile range (second row), root-mean-squared error (third row) and the proportion of patients who have an absolute percentage error more than 20% (P20) (fourth row) are displayed. All error bars are 95% confidence intervals calculated using bootstrap resampling with 2,000 repetitions and a normal distribution approximation.

quite arbitrarily but with the intention of minimising the errors caused by creatinine or GFR changing in the intervening time between the two measurements. The 30 days was suggested by clinical colleagues as reasonable. However, this should be checked as it is conceivable that

a patient's serum creatinine may change in the 30 days and thus affect the accuracy or bias of GFR estimated by any given model. Here we compare estimated GFR from the CamGFR model by the time between the nmGFR and serum creatinine measurements.

For the three datasets used in this section, the date difference between creatinine and GFR was not known for any of the patients in *Gothenburg* and a subset of patients in *Cambridge*. Figure 4.15 (A) shows the histogram of the difference in time between the GFR and creatinine measurement days. All patients within the *Manchester* dataset had their creatinine measurement on or in the week after their nmGFR. For patients from *Cambridge* the time difference was more variable. The mode for the time difference was zero days and the time difference was within a week for 81.4% of patients.

To compare the estimated GFR we then grouped patients by the time difference between their measurements. The groups for the time difference between serum creatinine and nmGFR measurements were as follows: more than one week before (n = 612), 3-7 days before (n = 714), 1-2 days before (n = 714), on the same day (n = 1,204), 1-2 days after (n = 822), 3-7 days after (n = 1,819) and more than one week after(n = 549). Patients for whom the time difference was not known formed another group called "NA" (n = 1,444). The accuracy (RMSE) and bias (residual median) of the refitted CamGFR for each of these groups is displayed in Figure 4.15 (B) and (C). We did not observe a decrease in accuracy as the time difference lengthened; if anything, we observed the opposite trend. Likewise, the patients whose creatinine was measured more than a week after their nmGFR, or more than a week before their nmGFR, were less biased compared with patients whose creatinine and nmGFR were measured closer together. The patients where the time difference was not known, had the highest RMSE and residual median. However, this is confounded by the dataset, as all patients from *Gothenburg* are included in this group. Other confounding variables might be involved in this analysis. In particular, patients who have a creatinine measurement close to their nmGFR may be more likely to be in-hospital patients which may affect the accuracy of estimated GFR. This is discussed in further detail in Section 6.1.

4.3.3 Discussion

In Section 4.2.1 we explained that the introduction of blood serum creatinine that is traceable to an IDMS reference material (IDMS-traceable) has improved the compatibility of creatinine across different centres but has also led to a decrease in the average creatinine value. This

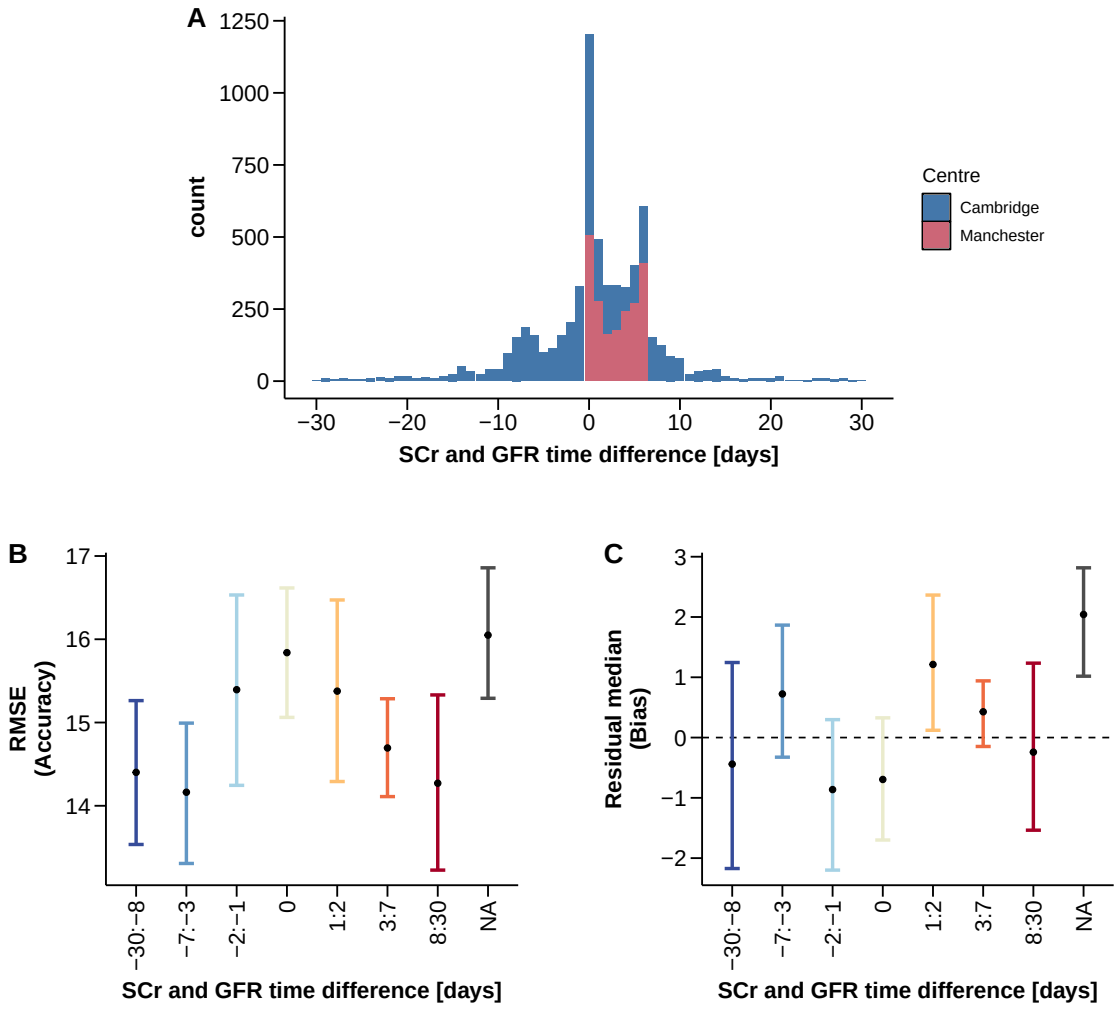


Figure 4.15 Comparison of CamGFR accuracy and bias as the time between serum creatinine (SCr) and GFR measurements increases. (A) Histogram of the SCr and GFR time difference, with bars coloured by the patients cancer centre. RMSE (B) and residual median (C) for the refitted CamGFR model when patients are grouped based on the SCr and GFR time difference. Negative time differences indicating that the serum creatinine was measured before the GFR. All patients from the *Cambridge*, *Manchester* and *Gothenburg* are included. Where the specific date was not known, these patients were included in the NA group. The number of patients in each group from left to right are: 612, 714, 531, 1204, 822, 1819, 549, and 1444.

decrease is probably attributable to the historical overestimation of creatinine when measured by the Jaffe method due to interference from other substances, in particular albumin. The CamGFR model was developed using data that were not IDMS-traceable and thus an adjustment was required for the CamGFR model to be suitable for use with IDMS-traceable creatinine. Hence, in this section, we refitted the previously published [60] and validated [132] CamGFR

model to be suitable for IDMS-traceable creatinine. This was achieved by adding an additional interaction term between the creatinine terms in the model and creatinine-type (IDMS or non-IDMS). We also explored alternative models (e.g. the piecewise linear and leave-one-out models) and these showed slight improvement in model accuracy. However this improvement was ultimately deemed not substantial enough to warrant changing the model form, which has the additional benefit in allowing clinical continuity. We thus now present the refitted CamGFR, which can be used to make a prediction for GFR based on either non-IDMS-traceable or IDMS-traceable creatinine data.

Using the multi-centre validation dataset we showed that the refitted CamGFR was the most accurate and least biased for both IDMS-traceable and non-IDMS-traceable creatinine data compared with previously published models. The improved accuracy of the CamGFR model in GFR estimation directly translates into an improved dosing accuracy for carboplatin dose calculations via the Calvert equation [13]. We further compared the CamGFR with other models for patients who did not have cancer where the model was equally as accurate.

Of the previously published model (excluding refitted CamGFR), Lund-Malmo was observed to be the most accurate model for IDMS-traceable data and CKD-EPI was the most accurate for non-IDMS-traceable data. This second observation is surprising given that CKD-EPI was developed using IDMS-traceable creatinine, yet using IDMS-traceable creatinine data in this thesis the model is less accurate and highly biased (with an overestimation of 10.58 ml/min) compared with non-IDMS-traceable creatinine data. Although we have insufficient information to provide a conclusive explanation for this result, one possible explanation could be that there are substantial differences in IDMS-traceable creatinine between institutes in different continents. Although this seems unlikely given the very purpose of IDMS-traceable creatinine is to reduce inter-centre differences, the CKD-EPI model was developed using a USA patient population while CamGFR and the Lund-Malmo were developed using European data. To provide a more conclusive explanation, additional data need to be collected from North American centres so as to compare the creatinine values and further generalise a model applicability.

The inclusion of interaction variables as candidate terms for the model enables the identification of more complex relationships between GFR and the explanatory variables. One particular example which was observed to be consistently significant in the modelling process was the BSA-age interaction term, implying that the effects of age and BSA on GFR are not merely additive. The fitted coefficient for this interaction term is negative, whilst the fitted coefficients

for age and BSA are positive and negative, respectively. Thus, one interpretation of the BSA-age interaction term is that as a patient's BSA increases, the negative association between GFR and age increases, as seen in Figure 4.13 D. Considering this interaction term from a biological perspective, and given that BSA is correlated with obesity, we note that this may reflect the contribution of microvascular disease to the loss of renal function [35]. Microvascular disease that, in turn, can worsen with age, hypertension and poorly controlled hyperglycaemia.

Despite improved accuracy, the refitted CamGFR model is still limited by a few important factors. First, although we had a few non-Caucasian patients in our study ($n = 22$), these were not enough to add an adjustment to the model. Previous models [77, 76] with a population of patients with different ethnicities have shown that an adjustment for certain ethnicities improves the model fit.

Second, although standardisation of serum creatinine assays has reduced the inter-centre variability, there are still underlying issues in using it as the primary renal filtration marker in a model for GFR. The primary issue is its dependence on non-kidney factors such as muscle mass, diet and supplements, as discussed in Section 2.4.2. Furthermore, disease processes or therapeutics may affect serum creatinine in an unknown and unpredictable manner which could lead to erroneous estimations for GFR. This has led to the suggestion of alternative filtration markers to serum creatinine, in particular cystatin C.

Recent literature has shown the potential benefit of using cystatin C as opposed to, or in addition to, creatinine [58, 89, 48, 101]. Cystatin C has an advantage over creatinine in that it is not as affected by muscle mass and diet, as a result models which use cystatin C do not typically require a race adjustment. However, cystatin C is not routinely measured, which is in itself a disadvantage of including cystatin C in a model. This was something evidenced in this thesis as the data were collected retrospectively and cystatin C had not been measured in standard patient blood tests and thus it was not measured for any of the patients in our study. A further disadvantage is that cystatin C has been shown to have similar issues to creatinine with regards to inter-centre variability. There is also some evidence for involvement of cystatin C in cancer progression, which might impact on its validity to be used as a filtration marker for patients with cancer [70]. Cystatin C is not the only other filtration marker to have been considered as an alternative to creatinine, other filtration markers have been included in previous models, namely urea and albumin. We explore these and other haematological and biological blood results in Chapter 5. These too have issues with inter-centre variation which impacts on their usefulness for estimation of GFR.

4.4 Conclusion

Having analysed data from 9,233 patients, we have developed a new model to estimate GFR. This work has primarily been focused on patients with cancer, for whom GFR is commonly estimated, but no recent model exists. This is important given that patients with cancer were typically seen to have normal GFR for their age, this is in contrast to chronic kidney disease patients who have often been used to develop previous models. We have shown that the refitted CamGFR model generally performs better than other published models, and, given further validation, should be considered as a useful estimate of GFR, particularly in patients with cancer.

It should always be remembered that any estimated GFR from a model has an associated variability, and thus there may be a large difference between the true GFR and estimated GFR for any particular patient. As such, clinicians should give careful consideration to the variability of any estimated GFR. In particular, when reporting an point estimate for GFR a confidence interval should also be provided. This is available for the CamGFR model via the online application published alongside the original manuscript [60]. However, wherever possible and particularly when a high level of confidence is required, GFR should be measured using nuclear medicine.

There are a number of questions and considerations that have arisen at the culmination of this chapter. Some of this we will investigate in the next chapters and also within future work, including:

- Whether for additional haematological and biochemical measurements, any correlation exists between these measurements and GFR, and the potential implication of this
- Whether inclusion of additional haematological and biochemical measurements will reduce the risk of inaccurate estimation and hence further improve the accuracy of a model to estimate GFR
- The possible adverse effects of chemotherapy treatment on a patient's GFR and serum creatinine, and the implications of this
- The issues with and implications of estimating a patient's GFR when their creatinine is not stable

- How renal function in patients with cancer compares with renal function for healthy kidney donor patients, and the effect of this on a model to estimate GFR

Chapter 5

Exploratory analysis of non-creatinine variables as predictors of GFR

The development of a model to estimate GFR in Chapter 4 focused on using creatinine as the only clinical measurement. However, in clinical practice many routine measurements are performed to assess a patient's general state of health, check for infection, or monitor organ function. In this chapter we assess whether these other routine clinical measurements may be predictive of GFR. We begin with exploratory analysis of these data, specifically examining association of these measurements with GFR. We then consider whether GFR can be more accurately estimated if these additional variables are included in a regression model. Finally, using the results of these models, we discuss the particular biochemical or haematological variables that are predictive of renal function.

5.1 Data description

The data that inform this analysis consist of a subset of *Cambridge* and *London-Barts* datasets used and described in the previous chapter. For patients from *Cambridge*, the data were collected by extracting all biochemical and haematological blood tests that were performed within 30 days of the given patient's nuclear medicine GFR measurement. This dataset consisted of 3,000 nuclear medicine GFR measurements (nmGFR) from 2,576 patients. For these patients, there were 39 biochemical and haematological blood test results, which included

serum creatinine, observed in 200 or more patients, with a total of 822,171 observations. Henceforth, the biochemical and haematological blood tests will be referred to more simply as blood tests.

To reduce the longitudinal data to single time point data, each of the blood tests was filtered, leaving only those closest in time to the nmGFR measurement. Results were further required to have been made within 30 days of the nmGFR measurement to be considered valid. This procedure meant that for a given patient, the different blood tests might not have been performed on the same day. In fact, only 44.9% of patients had all tests performed on the same day.

Not all blood tests were available for any given patient. The most commonly available blood tests were for creatinine, sodium, albumin, urea, potassium, alkaline phosphates (ALP), bilirubin, alanine aminotransferase (ALT), each of which was available for more than 1,900 patients. If the measurement units of the blood tests changed during the collection period, units were corrected where possible. However, this was not possible for Platelet Distribution Width (PDW) due to rounding errors, thus it was excluded.

5.2 Correlation of routine biochemical, haematological and demographic variables

We first examined the correlation between the continuous variables in the extended *Cambridge* dataset to study the strength of the relationships between variables. To do this, we used the Spearman correlation coefficient, as it captures monotonic relationships and not just linear relationships. Figure 5.1 displays the correlation matrix, in which variables are ordered by hierarchical clustering of the absolute value of the correlation coefficient.

Groups of correlated variables are observable in the correlation matrix and dendrogram. For example, white blood cell count and the immune cells eosinophils, basophils, neutrophils and monocytes are intercorrelated. Lymphocytes, platelets (PLT) and procalcitonin (PCT) are another intercorrelated group, which was also correlated with the previous group of variables. Similarly, measures associated with red blood cells were intercorrelated; this included red cell distribution width (RDW), red blood cell count (RBC), haematocrit and haemoglobin. As expected, height, weight and BSA, which is a function of height and weight, were intercorrelated.

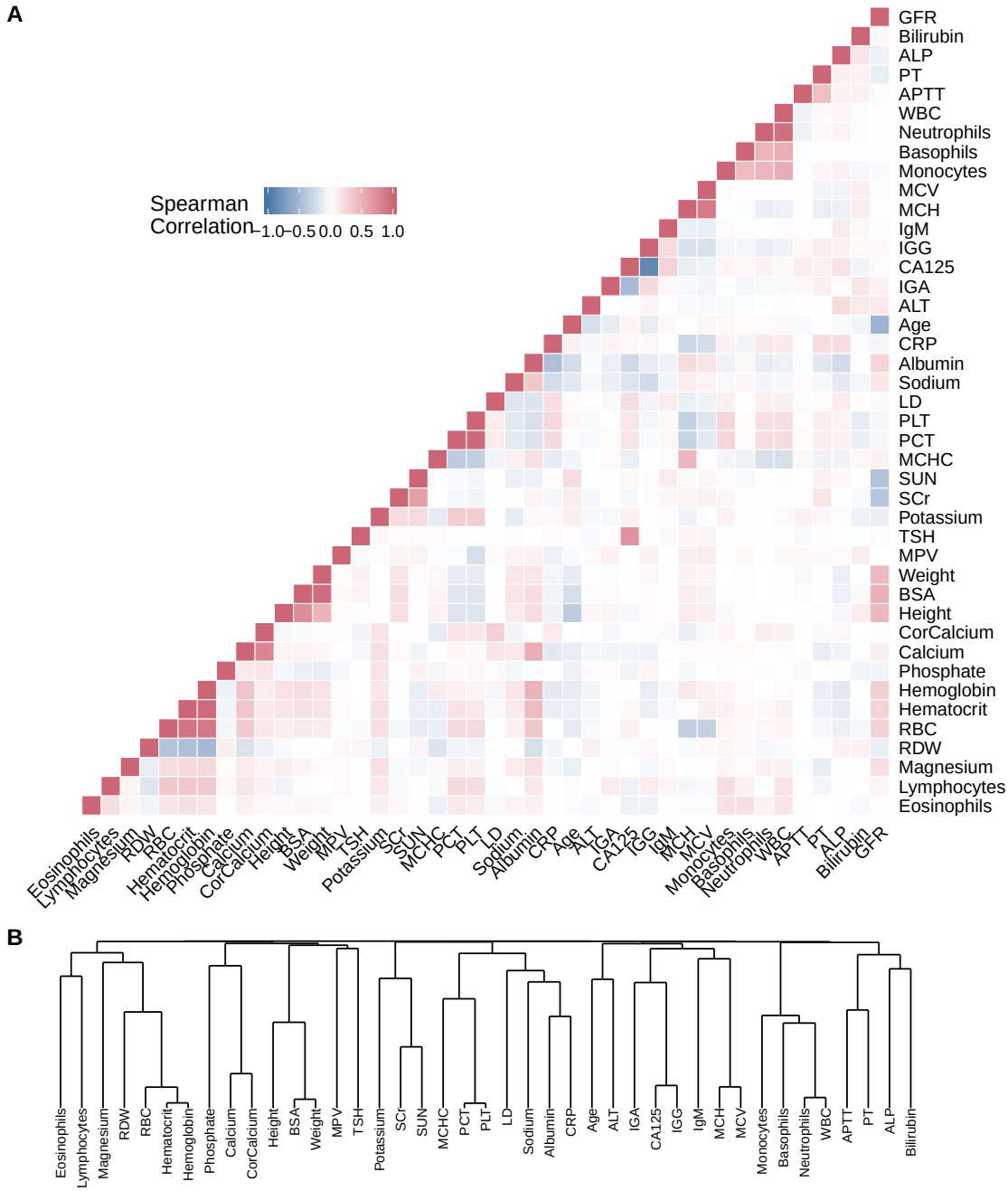


Figure 5.1 Correlation matrix of GFR along with the demographic , biochemical and haematological variables. (A) Spearman correlation matrix for the 42 continuous variables. Correlation was calculated for all complete pairs. With the exception of GFR, variables are ordered by hierarchical clustering with the distance between two variables x and y set to be $(1-|\text{cor}(x,y)|)$. The dendrogram for the hierarchical clustering is shown in (B).

When examining the correlation between GFR and all other variables, age, BSA, height and weight were the variables with the strongest correlation. All four had an absolute correlation higher than 0.4. Urea had a higher correlation with GFR than creatinine, -0.38 compared with 0.30 respectively. The other variables with an absolute correlation higher than 0.2 in order of magnitude were albumin, haemoglobin, haematocrit, RBC and ALT.

5.3 Regression modelling for non-creatinine variables

After examining the correlation between GFR and the covariates, we performed a regression analysis using the extended *Cambridge* dataset. This had two main aims, the first to examine whether GFR could be more accurately estimated by including additional blood test rests available in the extended *Cambridge* dataset. The second aim was to find variables which are associated with renal function after accounting for any confounding variables.

For the regression analysis we needed a complete dataset, which was not the case in the correlation analysis, to model the relationship between GFR and the predictive variables. Different imputation methods were considered but ultimately not used. Instead, a complete dataset was produced by excluding either patients or variables from the dataset. For this, we assessed the trade-off between preserving the number of different variables and preserving the sample size by examining the missing value pattern. Using this strategy, we excluded 13 variables with more than 1,468 (or 48.9%) missing values. Following this, the 1,490 samples with missing observations were excluded. The resulting dataset contained 1,510 samples from 1,361 patients and 29 variables including creatinine, age, sex, height, weight and BSA. Similar to the regression analysis in Chapter 4, this dataset was split into a model fitting (n = 1,208) and a model validation (n = 302) dataset at a ratio of 4:1.

For the subsequent analysis to be valid, we need to assume that the data are missing completely at random (MCAR). This assumes that the event that leads to a particular observation being missing is independent of both observed and missing data [51]. In reality, this is unlikely to be true for our data for several reasons. Blood test results are often measured together thus the absence of one is associated with the absence of another. Blood tests are more likely being obtained in periods of increased health care need due to acute illness or intense therapy. The value of one measurement may result in the other measurements being performed. For example,

variable	estimate
Intercept	4.425
BSA	2.825
Age	-0.023
Sex_M	0.577
Albumin	0.015
log(SUN)	-0.389
log(ALP)	-0.116
log(ALT)	0.080
Haemoglobin	0.004
RBC	0.161
log(RDW)	-0.144
log(SCr)	-2.503
$\log(\text{SCr})^2$	-0.278

Table 5.1 Non-zero coefficients and their estimated value from the lasso model. Variables are standardised before performing lasso regression.

increased markers of inflammation may lead a clinician to check more carefully for adequate liver function. Given the above caution should be observed in interpreting any findings.

The distribution of each continuous variable was then examined. As most of the variables were either counts or concentrations, many were positively skewed. This could have meant that the GFR variable relationship was not approximately linear. To help resolve this, we log-transformed all predictive variables whose skewness was greater than 1. A small adjustment of 0.1 was added to each variable, except for creatinine, before log transformation to resolve issues of log-transforming zero values. The variables which were transformed, by order of skewness, were APTT, bilirubin, PT, basophils, ALP, neutrophils, WBC, eosinophils, lymphocytes, creatinine, ALT, monocytes, CRP, urea, PLT, PCT, RDW, and MPV.

Compared with the regression analysis from Chapter 4, this analysis required us to choose a set of predictive variables from a more extensive list of candidate variables, many of which were likely to be uninformative. For this, along with the stepwise procedures described and used in Chapter 4, we also considered elastic net regression methods. Specifically, we used lasso, as this performs both regularisation and variable selection. Incidentally, it was also more accurate than other elastic net models. We did not consider pairwise interaction terms as they increased the number of variables to 781, the square and cubic log(SCr) terms were included.

Finally, we considered linear additive models (AM) to model the relationship between square-root transformed GFR and its predictive variables. This allowed us to model non-linear relationships between GFR and the predictive variables in a more general fashion. All continuous variables were modelled using P-splines [37]. Variable selection was performed for the additive model by including an additional penalty to a smooth term, allowing terms to be removed from the model. Once variables were removed, the model was refitted with the remaining coefficients. Further details on variable selection for additive models can be found in Section 4.1.2.

When performing stepwise selection with BIC as the selection criterion, the selected model contained the variables BSA and log(SCr), along with the squared and cubic terms, age, sex, haemoglobin, log(SUN), log(ALP), log(ALT), MCV, and albumin. All fitted coefficients were significantly different from zero.

Lasso regression was performed using 10-fold cross-validation to find the value of the penalisation parameter λ; see Section 3.2.2 for further details. The value which minimised the mean-squared error was $\lambda_{min} = 0.011$, and the value which was one standard deviation larger than this was $\lambda_{1se} = 0.057$. Using λ_{1se} the non-zero coefficients are provided in Table 5.1, along with their estimated value.

When fitting the additive model with our variable selection procedure, the non-zero variables were log(SCr), BSA, age, height, weight, albumin, log(SUN), potassium, log(ALP), log(ALT), RBC, log(RDW), log(monocytes), log(neutrophils), and log(lymphocytes). After removing the zero variables the model coefficients were re-estimated. All remaining coefficients significantly improved the fit of the model, with p-values calculated using an F-test for each variable. Figure 5.2 displays the fitted smoothed curves for each of the continuous variables in the model. Interestingly, the fitted curves for height and weight seem to indicate that as either variable increases, the predicted value of GFR decreases. This is contrary to the previously seen positive correlation between GFR and either variable (Figure 5.1). However, the plot is misleading in this instance; since BSA is calculated from height and weight, positive correlation between height or weight and GFR are absorbed into the effect of BSA on GFR. Hence, a given increase in height or weight would result in an increase in BSA, and subsequently an overall increase in estimated GFR. This case highlights the complexities of examining the individual effects of the variables, particularly for highly intercorrelated variables.

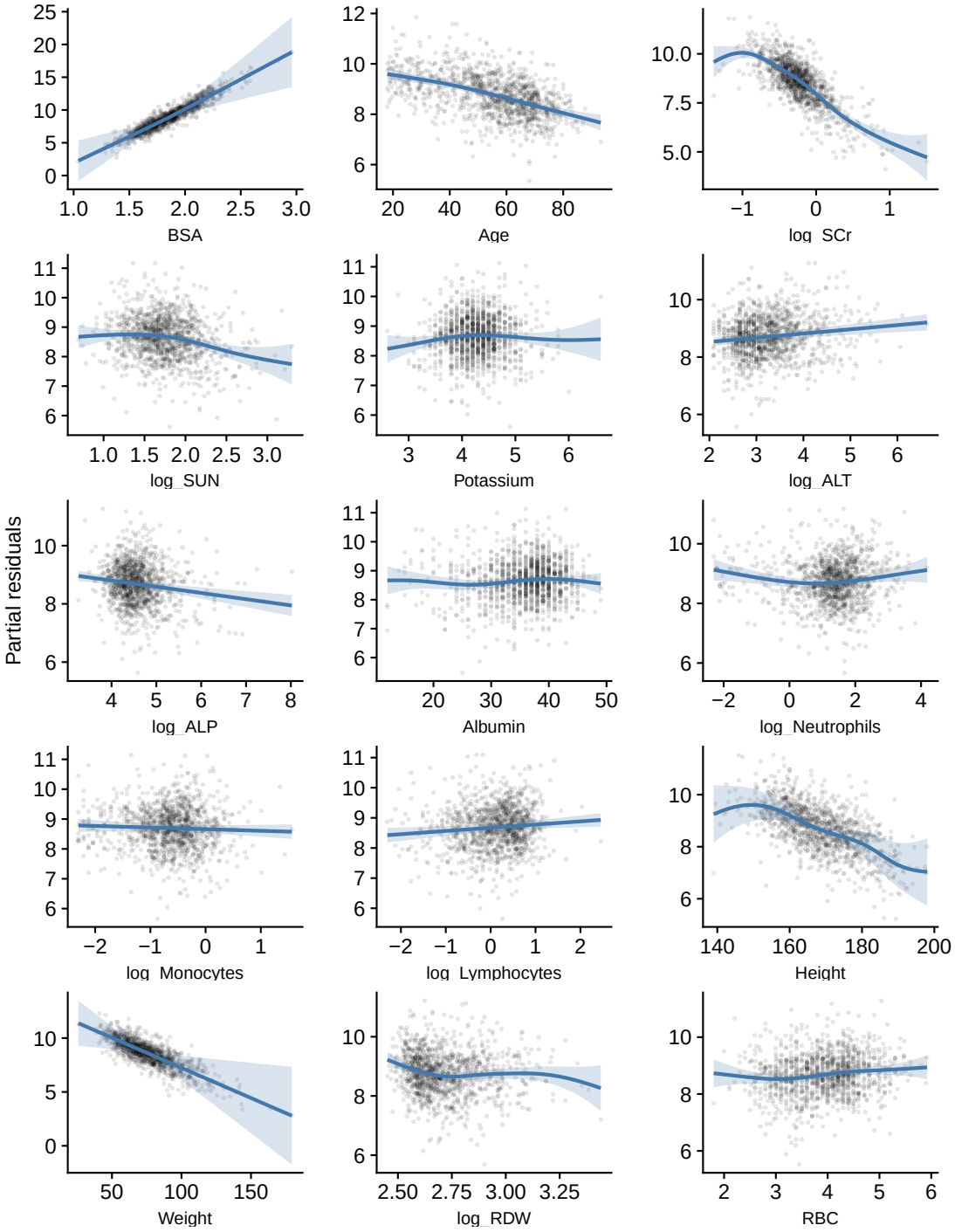


Figure 5.2 Fitted smooth functions for all variables in the additive model. Blue curves are the fitted smoothed functions for each continuous variable in the additive model. The functions are calculated conditionally on all other variables being equal to their mean and displayed on the square root scale. Background points are the partial residuals, which are equal to the sum of model residual and the fitted smoothed function for the given variable.

Three modelling approaches were considered in this section: stepwise selection with linear regression, lasso, and additive models. All of these models selected common variables, and all included the variables that were used for the refitted CamGFR model, specifically BSA, age, log(Scr) and sex. Variables for log(SUN), log(ALT), log(ALP) and albumin were also included in all three models. Finally, at least one variable associated with red blood cells was included in all of the models: haemoglobin for the stepwise model; haemoglobin, RBC and log(RDW) for the lasso model; and RBC and log(RDW) for the additive model.

5.3.1 Internal validation

To assess the performance of the new models, all three were compared with each other and with refitted CamGFR, CKD-EPI and Lund-Malmo on the separate validation dataset. Similarly to the approach in Chapter 4, models were compared using RMSE, residual median and residual IQR, which aim to assess model accuracy, bias and precision respectively. The more robust measure of accuracy P20 (the proportion of patients with an absolute error greater than 20%) was also used. All statistics were calculated using an estimated and observed GFR value on its ordinary scale to resolve the issue of different transformations used for the published models.

Table 5.2 provides the summary statistics for the six different models using the internal validation dataset. All three new models were more accurate than refitted CamGFR, which itself was the most accurate of the creatinine-based models. Of the three new models, the additive model was slightly more accurate than the other two, both as measured by RMSE and by P20. None of the new models, along with refitted CamGFR, had a residual median that was significantly different from zero. The patients that were included in these data overlap with the *Cambridge* dataset from Chapter 4 and all had their creatinine measured using an IDMS-traceable method. Hence, as previously seen, the CKD-EPI was very biased to overestimate GFR for these patients. In fact, over twice the number of patients had an absolute error greater than 20% when GFR was estimated using the CKD-EPI model, as compared with any of the three new models.

Overall, it appears that including the additional variables modestly improves the model. There is no evidence that any of the models are over-fitting the data. However, over-fitting may still be an issue given that the models have been trained and validated on patients from a single centre. Furthermore, similar to issues with inter-centre comparison of serum creatinine

model	RMSE	median residual	IQR	P20
AM	12.96 (11.70, 14.27)	-0.91 (-2.72, 0.79)	14.94 (12.36, 17.51)	0.175 (0.133, 0.218)
LASSO	13.23 (11.92, 14.58)	-0.56 (-1.85, 0.80)	16.36 (14.06, 19.13)	0.205 (0.161, 0.250)
Stepwise-BIC	13.23 (11.88, 14.63)	-0.70 (-2.22, 0.77)	15.19 (13.30, 17.23)	0.192 (0.148, 0.237)
CamGFR	14.02 (12.71, 15.37)	-0.70 (-2.02, 0.91)	17.24 (14.44, 20.38)	0.235 (0.188, 0.282)
Lund-Malmo	15.88 (14.40, 17.42)	-4.07 (-5.89, -2.05)	17.92 (15.26, 20.29)	0.295 (0.245, 0.344)
CKD-EPI	21.61 (19.72, 23.56)	-13.61 (-15.98, -10.75)	18.93 (16.48, 21.53)	0.464 (0.408, 0.520)

Table 5.2 Comparison of the additive (AM), LASSO, Stepwise-BIC with additional non-creatinine variables, with the creatinine-based refitted CamGFR, Lund-Malmo and CKD-EPI models. Models were compared by root-mean-squared error (RMSE), median residual, residual interquatile range (IQR) and proportion of samples with an absolute error greater than 20% (P20). All statistics were calculated using the 302 patients from the Cambridge internal validation data set containing the additional blood test results.

measurements, many of the new blood test results may be centre-dependent. The effects of any such dependence is currently unknown.

5.3.2 External validation

To try to resolve some of the issues raised in the internal validation we aimed to externally validate the model. In a similar situation to the development of the original CamGFR model, a small dataset composed of 108 male patients with germ cell cancer treated at Barts Health NHS Trust, London were used for this external validation. These patients had previously been used during the validation of the original CamGFR model for non-IDMS-traceable creatinine, Section 4.2.2.

There were some challenges in using these data for external validation, as some blood test variables that were available in the *Cambridge* dataset were absent from *London-Barts*. Importantly, the variables RDW, RBC, and monocytes were unavailable (these were included in the lasso or additive models). As a result, we were not able to validate directly the lasso and additive models but instead re-ran the model selection process using only the variables that were available in both datasets.

Rather than using external validation in order simply to validate previously developed models, here we use the *London-Barts* dataset to assess the impact of using a model developed at one centre on patients, or blood tests, measured at a different centre. It is particularly important to undertake this assessment because, unlike for creatinine, there is no international standardisation

for many of these blood test measurements. Hence, many of these measurements are likely to exhibit inter-centre variability [98], leading to possible bias and inaccuracy.

Before redeveloping the lasso and additive model, we first compared data from the two centres to assess differences in the variables. As all the patients from *London-Barts* were male and had a germ cell tumour, the same type of patients were extracted from the *Cambridge* dataset (n = 149). The difference in creatinine between these two centres was previously compared in Figure 4.7. Figure 5.3 shows the histogram for each continuous variable available in both datasets, with colour-coding to differentiate the two datasets. Both albumin and sodium immediately stand out as having differing distributions between the two centres, with higher values observed in *London-Barts* compared with *Cambridge*. This was also true to a lesser extent for log(ALP). The distributions of haemoglobin, log(lymphocytes), log(neutrophils) also all appear slightly different between the two centres. The values for eosinophils were more heavily rounded in *London-Barts*, leading to the observed differences in the histogram of log(eosinophils).

Some of these observed differences could be due to differences in the groups of patients at the two centres. Patients from the centres were matched by disease, however Addenbrooke's Hospital in *Cambridge* is a tertiary referral centre for germ cell tumours. Therefore, some of the patients in the *Cambridge* dataset may have had a more advanced tumour and may have been generally more unwell. Typically, we would expect a more unwell patient to have lower albumin, which is what we observe in the *Cambridge* dataset.

We did not know whether the creatinine values in *London-Barts* were IDMS-traceable or not. However, given that its distribution more closely resembled centres that had non-IDMS-traceable creatinine (Figure 4.7), we suspected that it was non-IDMS-traceable. The comparative creatinine data from the *Cambridge* dataset used in this chapter were all IDMS-traceable. The resulting difference in the distribution between the two centres can be observed in Figure 5.3, with the mode of the *London-Barts* to the right of *Cambridge*. However, the magnitude of this difference is far smaller than for the variables discussed above.

We then redeveloped the lasso and additive models using the *Cambridge* model development dataset, but considered only the variables that were available in both *Cambridge* and *London-Barts*. This reduced the number of variables for the development of the models from 29 to 19. For the lasso model this resulted in the variables RBC, log(RDW) being replaced by the variables log(lymphocytes) and MCV. In addition, the cubic log(SCr) term was now included.

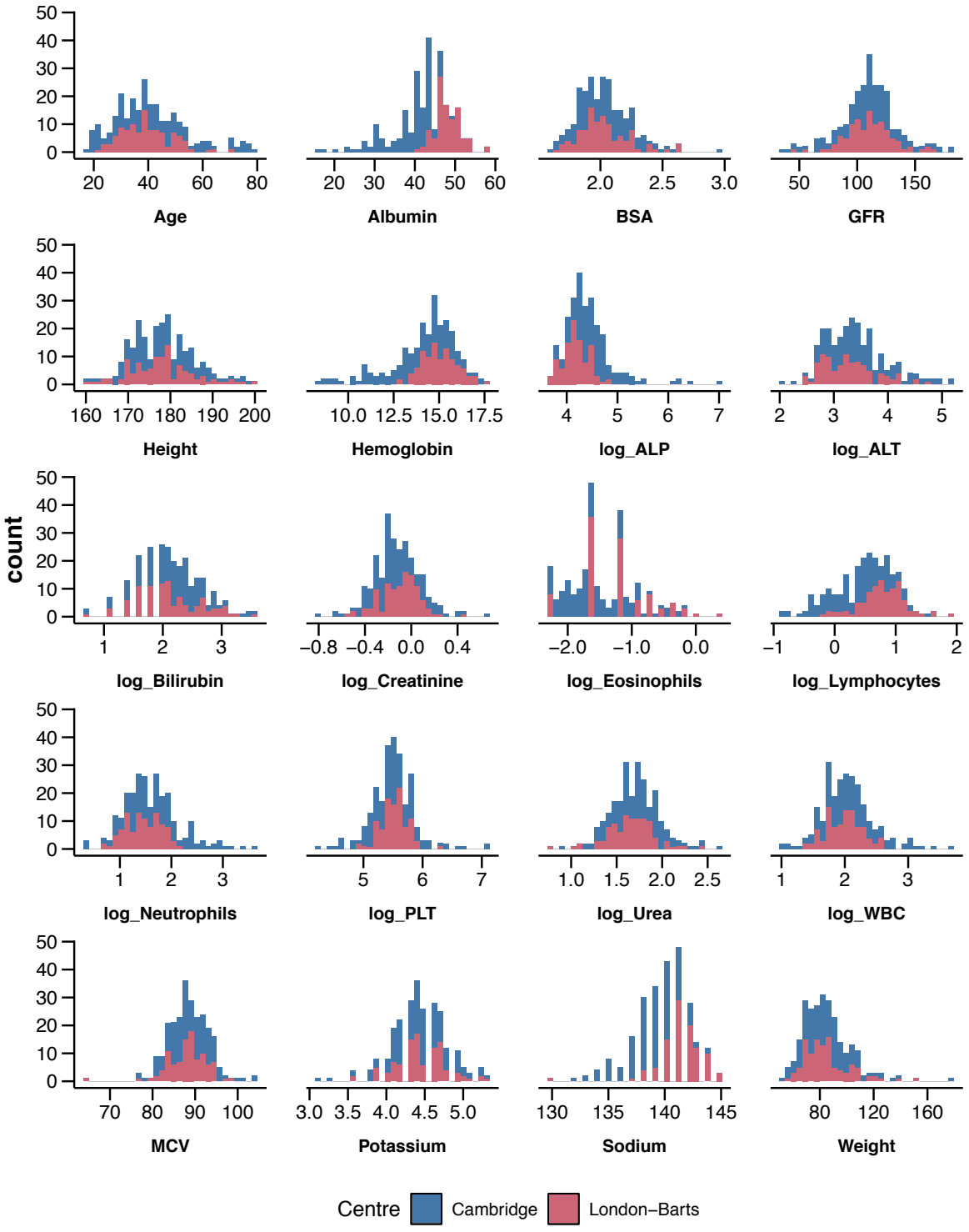


Figure 5.3 Histograms of all continuous variables included in both the *Cambridge* and *London-Barts* datasets. Patients were matched by diagnosis, with only the male germ cell patients (n = 149) included from the *Cambridge* dataset which was the only type of patient in the *London-Barts* dataset (n = 108). Variables are displayed on the scale used for modelling.

model	RMSE	median residual	IQR	P20
LASSO*	17.00 (14.45, 19.62)	1.98 (-1.08, 5.07)	22.82 (18.58, 29.22)	0.172 (0.095, 0.248)
AM*	17.28 (14.68, 19.94)	1.08 (-2.57, 5.59)	22.37 (18.88, 26.63)	0.172 (0.095, 0.247)
Stepwise-BIC	17.92 (15.11, 20.81)	-1.19 (-5.78, 3.46)	24.00 (20.28, 29.72)	0.182 (0.104, 0.259)
Lund-Malmo	17.98 (15.61, 20.43)	3.14 (-2.52, 7.88)	23.61 (19.56, 28.43)	0.202 (0.123, 0.280)
CamGFR-IDMS	19.14 (16.40, 21.96)	5.43 (0.89, 9.86)	22.29 (17.69, 27.11)	0.263 (0.174, 0.352)
CamGFR	19.94 (16.50, 23.55)	-5.92 (-11.49, -0.64)	22.23 (16.07, 26.09)	0.222 (0.141, 0.304)
CKD-EPI	20.81 (17.72, 24.07)	-11.06 (-16.62, -5.89)	24.72 (21.03, 30.73)	0.273 (0.184, 0.363)

* indicated that the model has changed from the model used for the internal validation

Table 5.3 Comparison of the additive (AM), lasso, stepwise-BIC (which contained additional non-creatinine variables), refitted CamGFR, Lund-Malmo and CKD-EPI models. Models were compared by root-mean-squared error (RMSE), median residual, residual interquatile range (IQR) and proportion of samples with an absolute error greater than 20% (P20). All statistics were calculated using the 99 patients from the *London-Barts* external validation dataset containing the additional blood test results. Creatinine data was annotated as being non-IDMS-traceable.

For the additive model the variables RBC, log(RDW), log(monocytes), height, and weight were replaced with log(eosinophils) and MCV. These changes did not result in the accuracy of the model changing when assessed using the internal validation dataset, with the lasso slightly more accurate than in its previous development. The RMSE (and confidence interval) was 13.19 (12.21 - 14.33) and 13.20 (12.22 - 14.34) for lasso and additive models respectively.

These two models were then compared with the BIC model, which did not need redevelopment, and with the creatinine based models (refitted CamGFR, LM and CKD-EPI) on the *London-Barts* dataset. Given that it was not known for sure whether the *London-Barts* creatinine values were IDMS-traceable or not, the GFR was estimated using the refitted CamGFR for both possibilities. Nine patients were excluded from the analysis as they had missing values for MCV, lymphocytes, eosinophils and SUN. This left a dataset that consisted of 99 patients to compare the models. Table 5.3 provides the four summary statistics used to compare the models using these patients. Examining the results we see that all three models that use the additional variables are more accurate than the creatinine-based models. All three new models along with the Lund-Malmo were unbiased, as evidenced by the confidence interval for the median residual containing zero. This was surprising given the observed difference in the distributions of the variables seen in Figure 5.3. Of the three models, the lasso was the most accurate, although the difference in accuracy between lasso and Lund-Malmo, which was the best of the creatinine-based models, was small.

5.4 Biological context of non-creatinine predictive variables of GFR

Thus far we have shown that including additional blood tests in models for GFR improves the accuracy of the prediction. Throughout the analysis, some blood tests were consistently observed to be predictive of GFR. In particular, SUN, albumin, ALP, ALT, and at least one red blood cell measurement (such as haemoglobin, RBC or RDW) were chosen in each of the models previously fitted in this section. Here, we will analyse the observed relationships between GFR and these four blood tests in more detail and discuss whether there may be biological causes for these relationships.

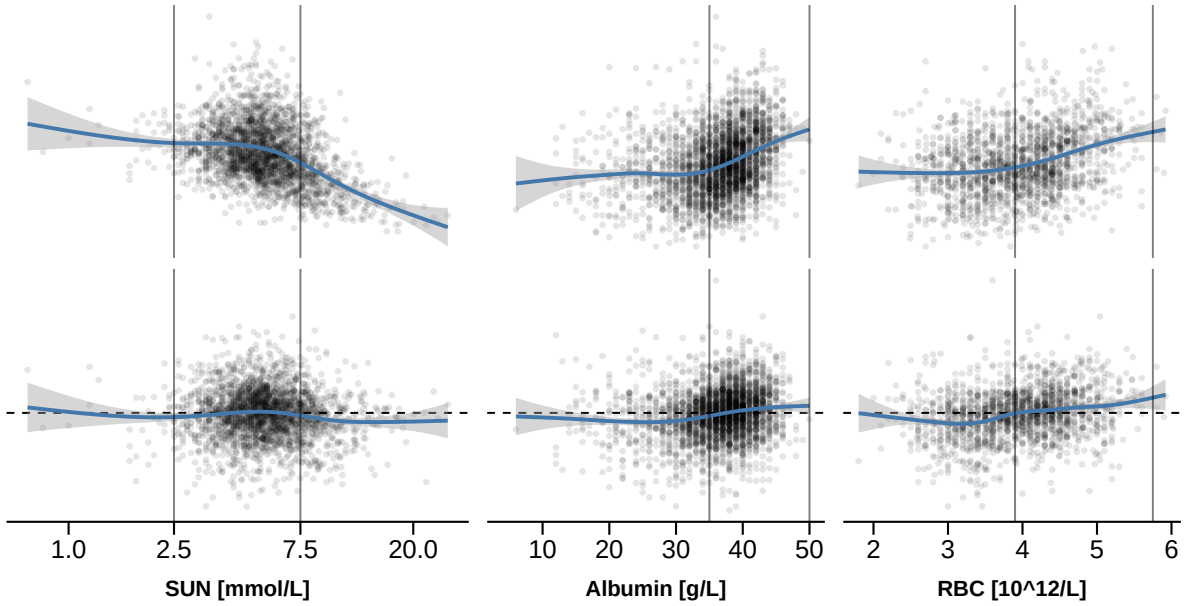


Figure 5.4 Serum urea nitrogen (SUN), albumin and red blood cell count (RBC) against either GFR (top row) or the residual when estimating GFR using the refitted CamGFR model (bottom row). SUN is shown on the log scale. The normal ranges for each variable are displayed as a solid vertical line. For RBC, the normal range depended on gender, and thus the maximum and minimum over both genders is shown. Smoothed lines are calculated using a generalised additive model with a cubic spline basis.

Figure 5.4 and 5.5 shows the relationship between both GFR and the residual of the refitted CamGFR model from Chapter 4 for SUN, albumin, one of the highly intercorrelated red blood cell measurements (RBC), ALP and ALT. In order to display the relationship between GFR and each blood test after the effects of creatinine and demographic data (used in refitted CamGFR) are removed, we have included plots of each blood test against the residual of the refitted

CamGFR. Additionally, to see the fitted effect of each variable in the additive model, Figure 5.2 can be examined.

First, we discuss serum urea nitrogen (SUN) and albumin, which are blood tests that have been used in previous models for GFR, specifically the six-variable MDRD [74]. As discussed in Section 2.4.2, SUN is a breakdown product in the nitrogen cycle and, similar to creatinine, it has been considered as an endogenous filtration marker and is functional-homologous to serum creatinine. In Figure 5.4, we see that the relationship between SUN and GFR is approximated by a piecewise continuous linear function with a spline point at approximately 4.0 mg/dl. For values below this point, as log(SUN) increases, GFR decreases. This is similar to the previously observed relationship between creatinine and GFR. Furthermore, SUN is the variable which has the highest correlation with creatinine ($\rho = 0.64$). It is therefore not surprising that, when examining the relationship with the residuals of the refitted CamGFR model, the observed association is reduced. In the additive model, the fitted function for SUN takes a similar shape to its raw relationship with GFR (Figure 5.2), indicating that SUN is still predictive even when both creatinine and SUN are included in a single model. In summary, SUN is acting as another measure of filtration function along with creatinine.

Serum albumin is a globular protein and is the main protein of human plasma. Its primary function is to regulate the oncotic pressure of blood. Albumin is produced in the liver, and medically is often used as a marker of liver function, with low concentrations a marker for liver disease. Additionally, reduction in serum albumin concentration may be caused by factors such as reduction in dietary protein intake, major surgery, systemic inflammatory disease, malignant conditions, or severe cases of the nephrotic syndrome [74]. Figure 5.4 shows the relationship between albumin and GFR. As with SUN, the relationship appears to be piecewise linear with a spline point at the the lower value of the normal range 35 g/L. Above this value, and for the normal range 35–50 g/L, there is a strong positive association between GFR and albumin. The positive association remains but is greatly reduced when examining the plot of albumin against the refitted CamGFR residual or in the fitted function for the additive model. This can be partly explained by examining its correlation with other variables included in the models, in particular: $\rho = 0.49$ with haemoglobin, $\rho = -0.25$ with ALP, $\rho = -0.24$ with age, and $\rho = 0.21$ with BSA.

Next we examine the relationship between GFR and red blood cell measurements: red blood count (RBC), haematocrit, haemoglobin, and red blood cell distribution width (RDW). RBC is the measure of the number of red blood cells per volume of blood, haematocrit is the percentage

of blood comprising red blood cells, and haemoglobin is an oxygen-transport metalloprotein found in red blood cells. RDW is a measure of the range of variation of red blood cells, which is negatively correlated with the previous three measures. Each of our models contained at least one of these measures and often more than one. Figure 5.4 shows the relationship between GFR and RBC. The plot would change little if haematocrit or haemoglobin were shown instead, given there is a strong positive correlation between RBC, haematocrit and haemoglobin (Figure 5.1). Examining this plot, for measured GFR, we observe that increased RBC is associated with lower GFR (Figure 5.4). Similarly to albumin and SUN discussed previously, the relationship appears to be piecewise linear with spline point near the lower normal range value 3.9×10^{12}/L. Above this point, and in the normal range 3.6–5.75 $\times 10^{12}$/L, there is strong positive association between GFR and RBC and below this point there is little association. The association remains after other covariates are accounted for, with a linear function fitted in the additive model (Figure 5.2). The additive model also fits a negative association for log(RDW), which itself appears piecewise linear.

There is a plausible biological interpretation for this observed relationship. Specifically, atrophy of the kidney may affect both its filtration function and other functions. One other function of the kidney is to produce a hormone erythropoietin which drives haematopoiesis, the production of the cellular components of the blood. Thus a reduction of erythropoietin may lead to a decline in red blood cell production, which would lead to a lower red blood cell count. So overall it is understandable that lower RBC is associated with lower GFR.

Finally, we examine alkaline phosphatase (ALP) and alanine transaminase (ALT), whose main use is as diagnostic tools in determining liver abnormalities. ALP is a homodimeric protein enzyme, which is involved in dephosphorylating compounds. It is present in all tissues, but is particularly concentrated in the liver, bile duct, kidney, bone, intestinal mucosa and placenta. ALP is used as a diagnostic tool in determining liver or bone abnormalities, in particular hepatitis, with very elevated levels being an indicator of tissue breakdown in the ducts of the liver. ALT is a transaminase enzyme involved in the alanine cycle, and is found in various body tissues but in particular the liver; significantly elevated levels of ALT are also known to be associated with liver issues [88, 79].

Both these blood tests are markers for liver abnormalities when elevated, however, the direction of the association with GFR for ALT and ALP is opposing. In the raw relationship (Figure 5.5 A, top), in the relationship with the residuals of refitted CamGFR (Figure 5.5 A, bottom) and in the additive model (Figure 5.2) we see increased ALP associated with decreased GFR, and

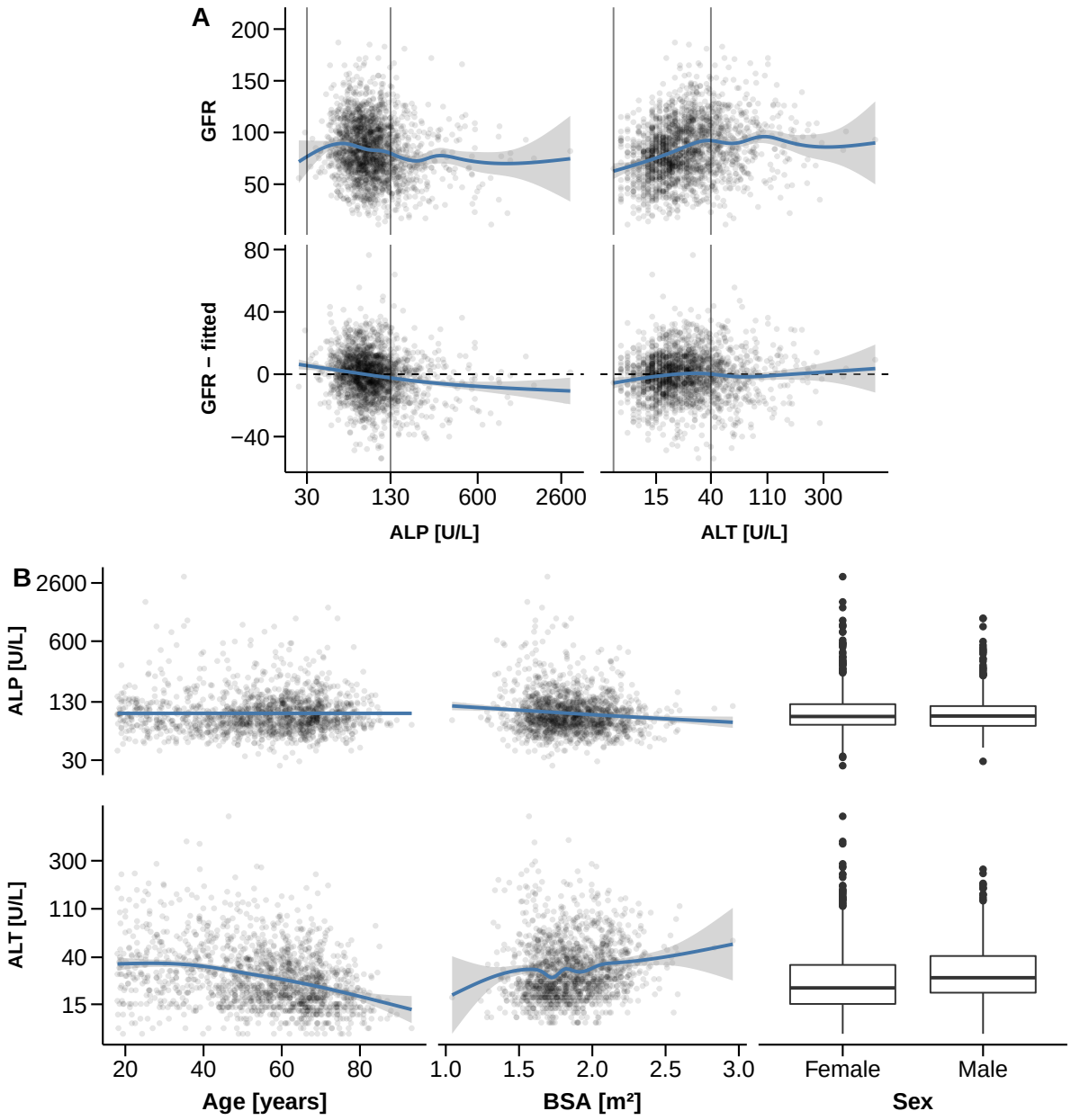


Figure 5.5 The relationship between Alkaline phosphatase (ALP) or alanine transaminase (ALT) and: GFR, the residual when estimating GFR using the refitted CamGFR model, age, BSA or sex. (A) ALP and ALT plotted against either GFR (top row) or the residual when estimating GFR using the refitted CamGFR model (bottom row). The normal ranges for each variable are displayed as a solid vertical line. (B) Relationship between ALT or ALP with age, BSA or sex, ALT and ALT ore shown on the log scale. For both plots ALP and ALT are shown on the log scale and smoothed lines are calculated a using generalised additive model with a cubic spline basis.

increased ALT associated with increased GFR. For ALP, the decrease in GFR is particularly seen for patients in the normal range of the measurement (30–130 U/L). Patients with an

ALP higher than the upper range of normal show less association, but have an average GFR of around 60 ml/min. When other covariates are accounted for, the decrease in GFR as ALP increases is seen more clearly throughout the range of ALP values and is more apparent than the relationship seen with ALT (Figure 5.5 A and Figure 5.2). For ALT, the increase in GFR is particularly apparent for patients within the normal range (7–40 U/L), with values higher than the upper limit showing less association. When other covariates are accounted for this positive association between ALT and GFR remains but is reduced (Figure 5.5 and 5.2).

Elevated levels of both ALP and ALT are associated with poor liver function, which may be associated with poor kidney function. It is accepted that at the extreme ends of the physiological spectrum such as cardiorenal and hepatorenal syndrome, as organ failure progresses, multiple organ failure may occur [91]. This relationship at the extreme of liver and kidney function may be a continuum of change and not a rapid absolute change. Indeed, in the above, we find that aspects that seem related to hepatic function become relevant when predicting renal function. Given this, the opposite direction of the ALT and ALP association with GFR appears to be contradictory. Both are markers for liver function when elevated, thus we would expect higher ALP or ALT levels to be associated with lower GFR yet this was only observed for ALP.

To explore ALP and ALT further, we looked at the relationship of both ALP and ALT, not just with GFR but also, with age, BSA and sex. This is because there are a number of generalisations that can be made based upon these demographics. In general, men have larger livers with greater liver function than women, people with a larger surface area have larger livers, and young people have better liver function than older people. Examining Figure 5.5 B, we see that ALT decreases as age increases, ALT increases as BSA increases, and there are higher values of ALT in men compared with women. For ALP, there is no association between ALP and age, no difference in ALP between men and women, and ALP slightly decreases as BSA increases.

These findings suggest that ALT within its normal range is a positive predictor of liver function. Although this is contrary to the usual interpretation of high ALT (multiples of the upper limit of normal) being associated with liver damage [79, 45], it is plausible for this to be the case if the liver tissue during normal turnover produces, in a mass-dependent manner, subtle leakage of ALT into the plasma. We could not find any literature to support or contradict this finding either way. Previous literature has recommended a higher normal range for men than women for ALT [44, 71], although the laboratory from which the data were obtained did not implement this. Furthermore, our dataset was not suitable for further examination as most patients in the dataset were early on in their cancer diagnosis and thus less likely to exhibit severe liver failure.

The negative association of ALP with renal function suggests that the tissue structure changes which lead to the release of the ALP molecule may always indicate early or progressing organ impairment. Again, though very high levels of ALP are known to be a marker for liver damage [79, 45], we did not have appropriate data to fully investigate this and could not find supporting or refuting literature for this subtle difference between ALT and ALP.

As a result of these inconclusive findings, we advocate caution in interpreting the possible relationships between liver function and GFR. Given the observed positive correlation between ALT and the demographics associated with higher GFR, the positive association between ALT and GFR may in part be driven by this. However, the positive association between ALT and GFR remains after these demographics are accounted for and is seen in the refitted CamGFR residual against ALT plot, in the fitted additive plot, and in the sign of the fitted coefficient for all models which included ALT. Therefore, even after the effects of these three demographics are accounted for there is still positive association between GFR and thus kidney function and ALT. Our findings justify follow up assessments for patients who are likely to have a wide spectrum of liver function.

5.5 Discussion and conclusion

In this exploratory study we have examined additional haematological and biochemical measurements and their association with GFR. In doing so, we have shown two key findings. First, that GFR may be more accurately estimated with addition of these variables and second, that in doing so, the observed relationships may lead to interesting biological observations.

We used three different regression methods to produce three different models, with little difference between them in accuracy. The three models were a linear model using stepwise selection to select the variables, a linear model with lasso penalisation, and a additive model (AM). There was a set of common variables in each of the models (age, SCr, sex, BSA, urea, albumin, ALP, ALT) and also variables that were unique to one or two of the three models. Dissimilar variable selection despite similar accuracy across the models could be interpreted in a number of different ways. It could be that variables specific to each model are not contributing significantly to model performance. Alternatively, given the inter-correlation between the additional variables, one individual variable from a group of intercorrelated variables may adequately explain the relationship with GFR. This later explanation is likely to be the case

for the intercorrelated set of variables that are associated with RBC, RDW, heamoglobin and haematocrit.

When comparing the models in the internal validation, the additive model was more accurate than the other two models and all three models were more accurate than previous creatinine-based models including the refitted CamGFR model developed in Chapter 4. In the external validation, the lasso model was more accurate than the other two and, again, all three models were more accurate than previous models, though the difference in accuracy between these three models and the Lund-Malmo model was small. The external validation provided further evidence that a model can be used to estimate GFR for a different centre from where it was developed without providing highly biased estimates, even when there are known systematic differences in the input variables.

If at this exploratory stage we were to suggest one of the three models to be used in order to estimate unobserved data, the lasso model would be preferred. In addition to being less complicated than the additive model, it is also likely to be the least over-fitted to the development dataset, as evidenced in the external validation.

Our next step in this analysis would be to collect these additional haematological and biochemical measurements from several different cancer centres and use them to compare measurements from different centres to assess their inter-centre variability. We would then seek to develop a model using data from across these centres that could be further internally and externally validated. Ideally, this model would be developed to be flexible enough not to need observations for all variables to estimate GFR.

A second key finding in this exploratory study was to show that a model developed using these additional variables can lead to interesting observations and hypotheses about the underlying biology. The resulting relationship between markers of liver function and GFR indicates possible links between a patient's liver function and kidney function. One could consider building models for integrated organ function using very large datasets based on these observations, however, this likely to be challenging in the case of other organs, which lack specific and measurable quantification of their function.

Chapter 6

Biological and clinical contexts of GFR estimation and measurements

In order to accurately estimate a patient's GFR, there are additional clinical considerations to be taken into account. In this chapter we discuss some of these in the context of the newly developed, refitted CamGFR model. First, Section 6.1 takes into account the potential adverse effects of chemotherapy on a patient's GFR and serum creatinine, and the subsequent impact on using models to estimate GFR for these patients. In Section 6.2 we examine the implication of using a creatinine-based model for patients with creatinine that is not stable by assessing the performance of the refitted CamGFR model using longitudinal data. Finally, in Section 6.3 we compare the similarities and difference between patients with and without cancer, with specific consideration of the impact of these scenarios on modelling GFR depending on the age of the patient. These further exploratory and statistical studies will draw into question the accuracy and validity of using refitted CamGFR to estimate GFR in particular clinical circumstances and given the availability of clinical data. It will moreover provide an insight into the direction of future work that will build upon the work in this thesis.

6.1 Estimating GFR in patients under active oncological management

This section aims to assess the effect of receiving chemotherapy on a patient's GFR and the subsequent implications this may have on using models to estimate GFR for these patients. To do so we would ideally like to have data that are annotated as to whether or not a patient was receiving treatment at the time of their nuclear medicine GFR measurement (nmGFR). We would then be able to compare these patients with those who received their nmGFR measurement prior to treatment commencing and assess the effect treatment has on GFR or creatinine. Unfortunately, these data were unavailable for patients within our datasets so, instead, two surrogate methods were used to categorise patients according to whether they were pre-chemotherapy, currently receiving chemotherapy or had received chemotherapy.

For the first surrogate method, patients who had multiple GFR measurements were compared. For these patients we assumed that the patients who had repeat measurements within a short period of time were likely to have had their second measurement whilst under active oncological management, and those who had a repeat measurement after a long period of time were likely to have had chemotherapy in the intervening time. For these patients, we assessed the change in GFR, creatinine and the residual of refitted CamGFR for the repeat GFR measurements.

For the second surrogate method, we compared patients who had numerous creatinine measurements in temporal proximity to their nmGFR with those who had fewer. Here, the main assumption was that patients who had more creatinine measurements were more likely to be unwell, more likely to be in-hospital patients, and more likely to be under active oncological management. Between these two groups of patients the bias and accuracy were compared for several models to estimate GFR, in particular the refitted CamGFR model which was developed in Chapter 4.

6.1.1 Patients with multiple GFR measurements

When developing the refitted CamGFR model in Section 4.3, repeat GFR measurements that were within a year of a patient's previous measurement were excluded. This exclusion criterion aimed to remove any nmGFR measurements that were measured while a patient was receiving treatment and also aimed to reduce the statistical complexities caused by repeat nmGFR

measurements not being independent. In this section, we further analyse data for patients who have repeat nmGFR measurements and explore any issues arising from using creatinine-based estimation models for patients who are receiving treatment.

In total, our dataset consisted of 696 patients with two or more nmGFR measurements. Patients who did not have cancer or whose diagnosis was not known (n = 36) were excluded, leaving 661 patients for this analysis. These patients were from three centers: *Cambridge* (n = 537), *Gothenburg* (n = 109) and *Edinburgh* (n = 15). Patients from *Cambridge* and *Edinburgh* had up to five repeat nmGFR measurements, while the patients for *Gothenburg* had up to 18. Sarcoma was the most common cancer type for patients from *Edinburgh* (n = 15) and *Gothenburg* (n = 41). These sarcoma patients, and patients from *Gothenburg* generally, had short periods separating their repeat nmGFR, with the median time between a patient's first and second measurement being 83 and 93.5 days for *Edinburgh* and *Gothenburg* respectively. This contrasts with *Cambridge* where patients with gynecological cancer (n = 205) were the most common, and for whom the median time between first and second nmGFR was 602.5 days. Overall, the median period between the first and second nmGFR measurement for patients from *Cambridge* was 537 days.

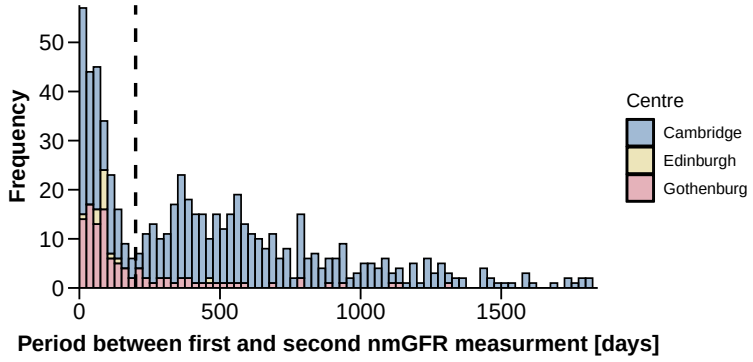


Figure 6.1 Histogram of the period between the first and second nmGFR. Each bar is 25 days wide and coloured by the centre. A dashed vertical line is displayed at 200 days. Patients with a period larger than five years are excluded for display purposes (n = 13).

These differences likely reflect the two common reasons for a patient with cancer to have repeated nmGFR measurements. The patient could have further measurements under active oncological management to assess whether the dosage should be adjusted; thus, there are short periods between nmGFR. Alternatively, the patient has one nmGFR at the beginning of platinum-based chemotherapy, with repeat nmGFR performed if the patient has relapsed before

beginning another chemotherapy regimen. These two groups of patients can be distinguished by examining the histogram of the period between the first and second nmGFR, Figure 6.1. The histogram is bimodal with a trough around 200 days where no patients were observed with a time period between 185 to 205 days. Henceforth, subsequent nmGFR with a period less than 200 from the first nmGFR will be referred to as short-period, and nmGFR with a period greater than 200 days will be referred to as long-period.

We first examined the changes in GFR for repeat nmGFR. Between the first and second nmGFR measurement, GFR significantly decreased by an average of 6.35 mL/min (median 5 mL/min). This decrease can be seen when the change in GFR is plotted against the difference in time between the first and all subsequent nmGFR, Figure 6.2 (A). In the figure, we observe the change in GFR is small when the time between nmGFR is short, with an average decrease of 0.526 mL/min for nmGFR within 50 days of the first nmGFR. However, the decrease is significant for both the short-period and long-period nmGFR, with a mean reduction of 4.44 and 7.40 mL/min respectively.

The observed reduction in GFR is not matched by an increase in log(SCr), the main variable likely to change in any creatinine-based model for GFR in the short term, Figure 6.2 (B). While there is a significant increase in creatinine for the long-period nmGFR measurements (mean increase 0.054), there was no significant increase in creatinine for the short-period nmGFR (mean increase 0.002). The other variable that may fluctuate in this period of time is weight. In the short time-period, no significant change in weight was observed (mean change -0.05 kg, p-value 0.067), which is in contrast to the significant increase in weight seen in the long-period (mean change 1.8 kg, p-value <0.0001).

Given that GFR decreased with a corresponding increase in log(SCr) and with only a slight decrease in weight, it is likely that the bias of creatinine-based models will be affected. Given that the data contain both IDMS-traceable and non-IDMS-traceable creatinine, only the refitted CamGFR model was examined as this should be unbiased for both. When GFR is estimated using the refitted CamGFR model there is a significant bias to underestimate GFR, particularly for the short-period nmGFR, Figure 6.2 (C). The first nmGFR measurement for each patient was unbiased with a median residual 0.427 mL/min (95% CI -0.547 to 1.26). The median residual subsequently reduces to -4.33 mL/min (95% CI -6.54 to -1.62) for short-period nmGFR, which is a significant bias to overestimate GFR; this bias is then reduced but and insignificant for the long-period nmGFR with a median residual -1.45 mL/min (95% CI -2.92

to 0.879). Consequently, the accuracy of the model is affected, with a RMSE of 15.4, 16.6 and 13.7, for the first nmGFR, short-period nmGFR, and long-period respectively.

Further examining Figure 6.2 (A), although overall GFR decreases in the short-term, for a few patients a large increase in GFR is observed. There were five patients who had a repeat measurement within two months and for whom their GFR increased by more than 40 mL/min in the second measurement. For these patients it is possible that they originally had an obstruction to the renal system, caused by or in conjunction with the cancer diagnosis, which was subsequently cleared before starting treatment. Alternatively, the clinician may have requested a repeat measurement as they did not trust the accuracy of the original measurement. For example, if a patient was dehydrated for the original measurement and thus it was necessary to repeat the nmGFR measurement [121]. Examining the residuals of the refitted CamGFR for these patients, they all have an estimated GFR which was at least 18 mL/min larger that the measured GFR, which suggests that indeed the measured GFR was inaccurate for the original measurement.

This observed reduction in GFR without an increase in creatinine is particularly apparent for the patients from *Edinburgh*. All these patients were diagnosed with a sarcoma, in particular either osteosarcoma, Ewing's sarcoma or desmoplastic small-round-cell tumor, and were aged under 50 (median 28 years). Figure 6.3 shows the measured and estimated GFR for the eight patients from *Edinburgh* who had at least three nmGFR measurements. Along with displaying the refitted CamGFR, the CKD-EPI model is also provided as this has been shown to be the most accurate and is unbiased for non-IDMS-traceable creatinine data (Figure 4.14). For seven of these eight patients, the nmGFR decreased at least once without a matching decrease in the estimated GFR. For the one other patient (patient ID: Edinburgh 341), their nmGFR increased in the two subsequent measurements with the estimated GFR increasing and then slightly decreasing.

In summary, it appears that patients who have repeated nmGFR measurements within a short period of time have a decreased GFR without a matching increase in creatinine. This leads to overestimation of GFR using creatinine-based models. The most likely explanation for repeat nmGFR measurements in a relatively short period of time is for the first measurement having been performed before starting treatment and for subsequent measurements performed whilst under active oncological management. Other possibilities for repeat nmGFR, such as obstruction to the renal system, are possible but would result in an increase to GFR. Patients

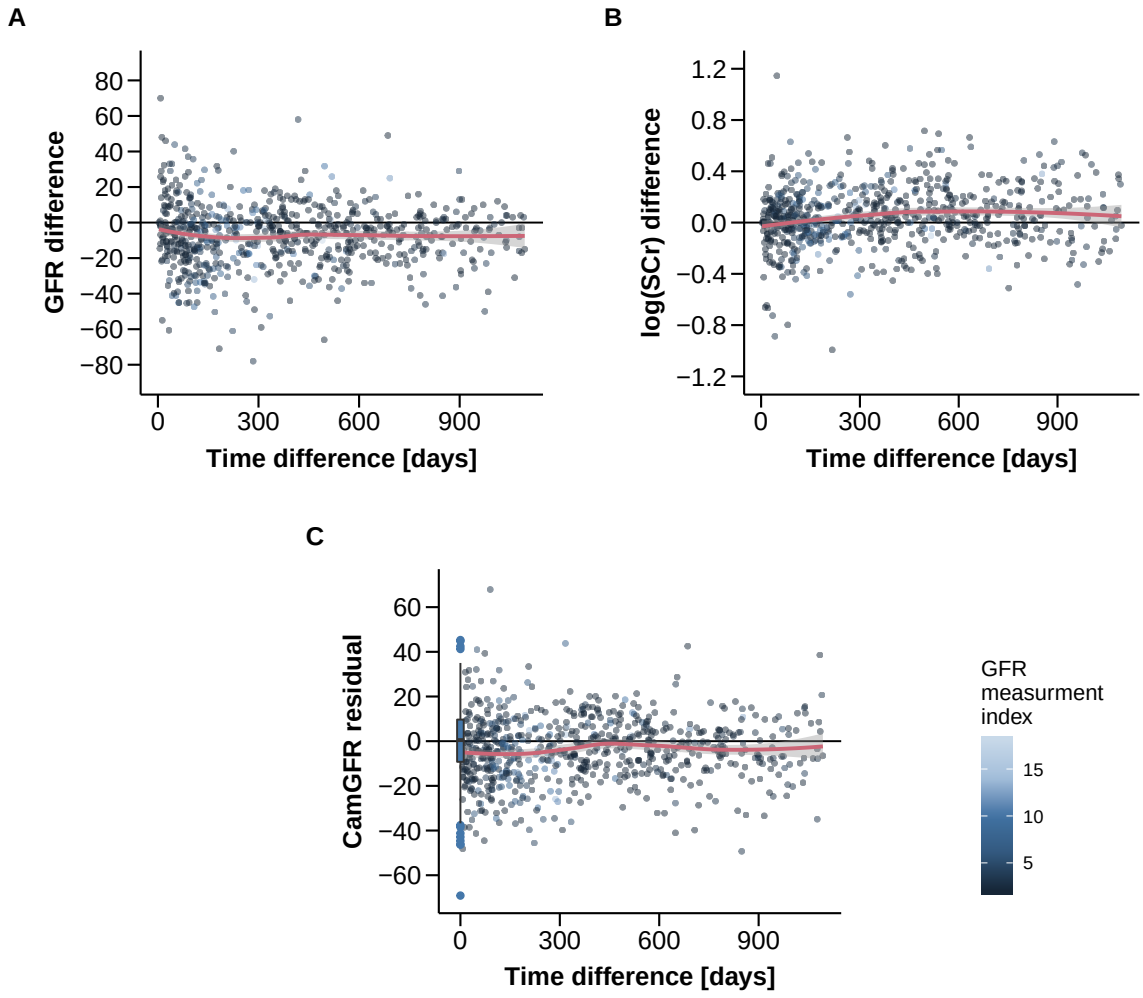


Figure 6.2 The difference in GFR (A), the difference in log(SCr) (B), or the residual of the refitted CamGFR model (C) is plotted against the time difference between the first and subsequent nuclear medicine GFR measurements for patients with at least two measurements within three years. Points are coloured by the number of previous measurements for the given patient. All data smoothing lines are loess curves. Repeat measurements with a gap of more than three years are excluded. A boxplot is displayed for the refitted CamGFR model residual for the first nmGFR for each patient (C).

with a repeat nmGFR measurement over a longer period of time also have a decreased GFR, however, these patients also have a matching increased creatinine.

These findings are corroborated by a previous study examining the post-chemotherapy effect on patient GFR and serum creatinine. Lauritsen et al. studied 390 germ cell patients who were treated at a Department of Oncology, Rigshospitalet, Copenhagen, Denmark between 1984 and 2007 [73]. They compared a patient's measured GFR, serum creatinine and estimated GFR before, after 1 year, after 3 years and after 5 years of receiving chemotherapy (bleomycin,

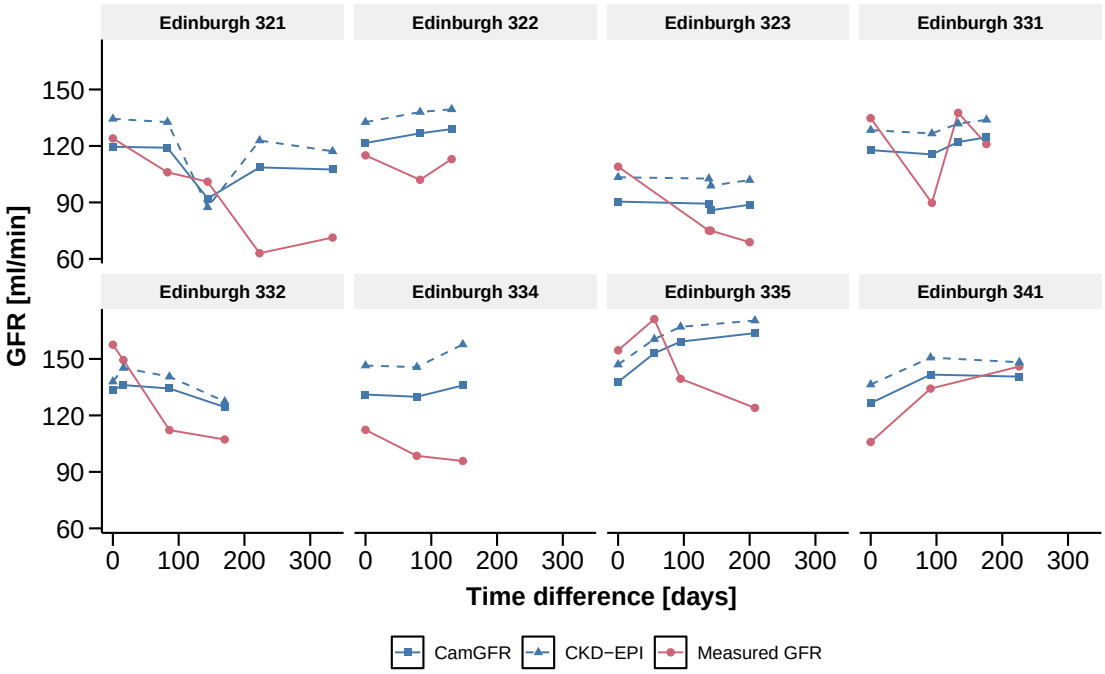


Figure 6.3 Measured GFR and estimated GFR, using either the refitted CamGFR model or CKD-EPI model, for eight patients diagnosed with a sarcoma from *Edinburgh*. Patients were included if they had at least three nuclear medicine GFR measurements (nmGFR). Time difference is the time between a patient's first nmGFR and subsequent nmGFR. Patients are named via an anonymous patient ID.

etoposide and cisplatin). In this publication, both measured GFR and creatinine were observed to significantly decrease compared with pre-chemotherapy levels. This led to the examined models for GFR (CKD-EPI, Wright, MDRD, Cockroft-Gault) generally becoming more biased to overestimate measured GFR, the exception being MDRD which was biased to underestimate GFR prior to chemotherapy. Furthermore, GFR that was measured one, three or five years after receiving chemotherapy were observed to have recovered, although not to the pre-chemotherapy level.

One weakness of the Lauritsen study is that it presents the change in GFR and creatinine on a group level and not at an individual level. This is further compounded by the fact that not all patients who had an initial GFR measurement before treatment began, had subsequent follow up measurements. This means that a confounding effect cannot be ruled out. For example, the observed decrease in GFR could be explained if there were patients who had measurements

taken before treatment but did not have a subsequent measurement, where their original GFR measurement was high compared with the group overall.

6.1.2 Patients with multiple creatinine measurements

As mentioned in the introduction of this chapter, the second method used to categorise patients was by the number of creatinine measurements in temporal proximity to their nmGFR. For a subset of patients from *Cambridge*, in addition to having the serum creatinine measurement that was taken closest to the nmGFR measurement, all serum creatinine measurements performed in the temporal vicinity of their nmGFR measurement were also obtained. Typically, the serum creatinine measurements were within 30 days either side of the nmGFR measurement. Using these data we aimed to further corroborate some of the findings in the previous section. Specifically, we assumed that if a given patient had numerous creatinine measurements prior to their nmGFR measurement this meant the patient was likely to be an in-hospital patient or a patient currently receiving chemotherapy treatment. Thus we were examining more than the effect of chemotherapy, and were looking more generally at the effect of unwellness on model performance. To assess the performance of creatinine-based models, patients were grouped by the number of creatinine measurements in the two weeks prior to their nmGFR, the accuracy and bias of GFR estimated using either Lund-Malmo, refitted CamGFR or lasso (developed in Chapter 5) was then compared for these groups.

For this comparison we used the same patients that were used to develop and validate the lasso model. These were chosen to allow the lasso model to be compared with refitted CamGFR and Lund-Malmo. The Lund-Malmo model was chosen as it was the most accurate for IDMS-traceable creatinine data (Figure 4.14), which all these patients had apart from for refitted CamGFR. Again, as in Section 6.1.1, non-cancer or patients with unknown diagnosis were excluded (n = 119). The remaining 1,391 patients were split into three groups based on how many creatinine measurements they had in the two weeks before their nmGFR (n_{SCr}). The groups were as follows: patients who had n_{SCr} between 0 and 3 (n = 1,213), 4 and 7 (n = 111), and 8 and 22 (n = 67). These cutoff values were chosen to discriminate between patients who were likely to be out-patients and not currently receiving treatment, out-patients who were receiving treatment, and in-patients. These groupings were not expected to be entirely accurate but to provide a gauge as to where there was an effect on the accuracy of the creatinine-based models. Indeed, when we examined the composition of the two groups with n_{SCr} equal to four

or more, we saw that these groups were enriched for patients with a haematological cancer. In the three groups 452 (37%), 80 (72%) and 52 (78%) patients had a haematological cancer diagnosis in each of the groups respectively.

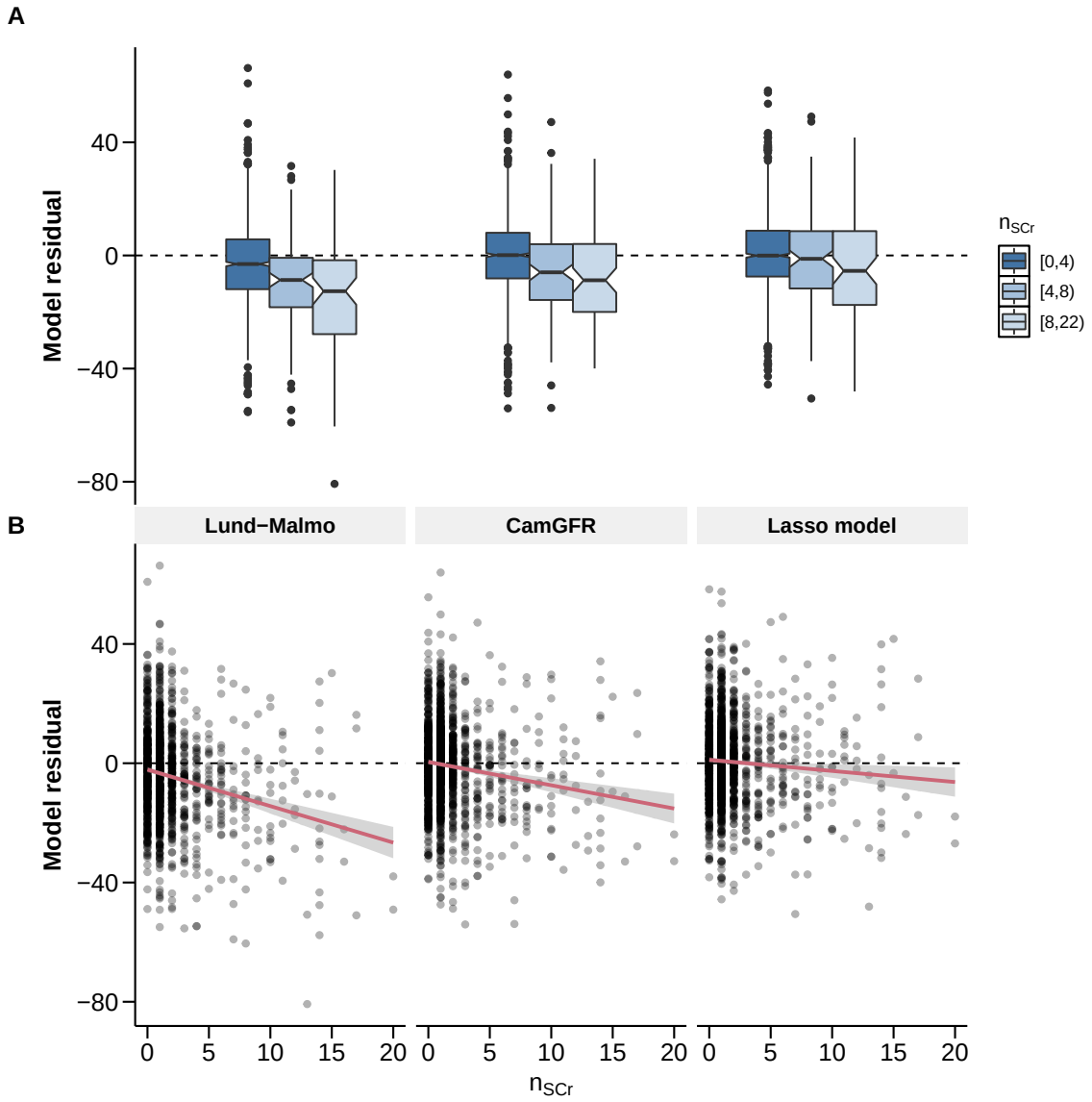


Figure 6.4 Boxplot and scatter plot of residuals for the Lund-Malmo, refitted CamGFR and lasso models against the number of creatinine measurements in the two weeks prior to the nmGFR measurement (n_{SCr}). (A) Boxplot of each model for patients groups by whether n_{SCr} was between 1 and 3 (n = 1,320), or 4 and 7 (n = 117), or 8 and 22 (n = 73). Notches are drawn at $\pm 1.58\,\mathrm{IQR}/\sqrt{n}$, and provide an indication of whether the medians differ. (B) Model residual against n_{SCr}. The line drawn was calculated using linear regression. The gradient of the line is significantly below zero for the Lund-Malmo, refitted CamGFR models, (p-value <0.0001 for both), but not for lasso model (p-value 0.095)

n_{SCr}	n	RMSE	median residual	IQR	P20
Lund-Malmo					
0-3	1213	14.73 (14.01, 15.45)	-3.09 (-4.00, -2.21)	17.28 (16.16, 18.38)	0.261 (0.236, 0.286)
4-7	111	18.38 (15.72, 21.15)	-8.31 (-11.58, -4.80)	17.86 (13.63, 22.10)	0.324 (0.237, 0.410)
8-21	67	26.34 (21.41, 31.32)	-12.24 (-17.07, -6.26)	26.91 (17.07, 35.29)	0.448 (0.329, 0.564)
refitted CamGFR					
0-3	1213	13.73 (13.03, 14.43)	0.20 (-0.68, 1.01)	15.97 (14.65, 17.35)	0.202 (0.179, 0.224)
4-7	111	16.54 (13.95, 19.25)	-5.27 (-10.41, -1.69)	19.40 (16.16, 23.28)	0.342 (0.255, 0.428)
8-21	67	18.11 (15.99, 20.33)	-7.34 (-11.54, -3.59)	24.08 (15.93, 32.35)	0.388 (0.275, 0.503)
Lasso model					
0-3	1213	13.32 (12.66, 14.00)	0.28 (-0.73, 1.12)	15.72 (14.54, 16.75)	0.186 (0.164, 0.208)
4-7	111	15.62 (13.05, 18.33)	-0.45 (-3.24, 2.58)	18.98 (15.68, 22.34)	0.225 (0.151, 0.301)
8-21	67	18.82 (15.85, 21.92)	-1.91 (-6.52, 3.42)	23.33 (15.47, 31.61)	0.388 (0.275, 0.502)

Table 6.1 Comparison of GFR estimated using either the refitted CamGFR, lasso or Lund-Malmo for patients split by the number of creatinine measurements in the two weeks prior to their nmGFR (n_{SCr}). The root-mean-squared error (RMSE), median residual and residual interquartile range (IQR) are given along with 95% confidence intervals.

Figure 6.4 shows the trend of the residuals for each patient group for the Lund-Malmo model, refitted CamGFR model, and lasso model as n_{Scr} increases. From this figure we observe the median residual reduces as n_{Scr} increases for all three models. If a linear regression is fitted for the residual against n_{SCr} the fitted slope of this regression is significant below zero for the Lund-Malmo, refitted CamGFR models, (p-value <0.0001 for both), but not for lasso model (p-value 0.095). When comparing the residuals between the three defined groups we observe that the refitted CamGFR is unbiased for patients who have n_{Scr} between 0 and 3 but has a significant bias to overestimate GFR for n_{Scr} more than 3, with a residual median of -8.41 mm/min for n_{Scr} between 8 and 22. The Lund-Malmo model is biased to overestimate GFR in all three groups. Finally, the lasso model was the only model to not have a significant slope for the model residuals as n_{Scr} increased. In addition, the residual median was not significantly different from zero for any of the three groups for the lasso model.

The observed increase in bias for each model as n_{SCr} increased was accompanied by a decrease in model precision and accuracy, as displayed in Table 6.1. The RMSE, the residual IQR and proportion of patients with an absolute error of more than 20% (P20) all increase as the n_{SCr} increase for all models. In particular, P20 was almost double for the patients with n_{SCr} between 8 and 22 compared with n_{SCr} between 0 and 3. These differences between the groups for all statistics were significant, as evidenced by the non-overlapping confidence intervals.

Overall, as with Section 6.1.1, there is a link between patients who are more likely to be unwell and overestimation of creatinine-based models for GFR. This overestimation is both significant and is associated with less accurate GFR estimations using any of the tested models. Of the three models tested, the lasso model was least biased for patients who had a high n_{SCr}. This is an indication that a model which includes other blood test measurements may negate the overestimation of serum creatinine only models.

There are several limitations that need to be considered when discussing these results. First, our assumption that patients who had a higher n_{SCr} were more unwell is not completely specific. There may be several other reasons why a patient would received numerous serum creatinine measurements. One such example is the potential inter-departmental bias for the number of serum creatinine blood tests performed, where some departments perform more blood tests than others. This effect is observed in the data with the high proportion of haematological patients in the group with the highest n_{SCr}.

6.1.3 Discussion

In Section 6.1.1 we showed that patients who are assumed to be under active oncological management, in particular whilst receiving or having received platinum-based chemotherapy, have their GFR underestimated when using a creatinine-based model. This underestimate is the result of a reduction in GFR, likely caused by the treatment, without a corresponding increase in serum creatinine. An implication of this is that any model that uses creatinine as its only filtration marker will not pick up loss of renal function. After a long time period has elapsed since the start of treatment, GFR remains stable and serum creatinine increases, models to estimate GFR, in particular refitted CamGFR, are then no longer biased to overestimate GFR. This finding is particularly relevant to clinical medicine where estimated GFR is used in order to adjust the dose of carboplatin whilst a patient is undergoing treatment: a large overestimate of GFR could lead to patient harm in the form of increased and preventable toxicity [65].

As previously discussed, it would appear that serum creatinine remains stable independent of changes to GFR both during and after chemotherapy treatment. There are a few possible reasons for this. The value of serum creatinine is known to be dependent upon factors such as muscle mass, physical activity and, to a lesser extent, it is also dependent upon dietary intake of protein and other nutrients. Many chemotherapy drugs are thought to cause sarcopenia [104, 109], which a decrease in serum creatinine is a marker for [96]. Indeed, we observe a

small decrease in weight between subsequent measurements of GFR within our dataset. The other influencing factors might too cause a reduction in serum creatinine, and so any increase that would correspond with reduced GFR could be negated.

Findings from Section 6.1.2 compliment the above, though in this second section patients were categorised slightly differently dependent on the number of serum creatinine measurements in the two weeks prior to the nmGFR measurement (n_{SCr}). Thus patients were not necessarily receiving, or had just received, chemotherapy treatment but were likely to be unwell with something that was related to their cancer diagnosis. Similar to the previous section, patients who had a high n_{SCr}, and thus were assumed to be unwell, were overestimated by the three creatinine-based models assessed. It is likely that, as we observed in Section 6.1.1, these patients had a reduced GFR without a corresponding increase in serum creatinine, thus leading to the overestimation.

The model that shows the least amount of overestimation, and thus is most suitable for these unwell patients, was the lasso model developed in Chapter 5. This is particularly interesting as it was the only model to include other biological and haematological measurements in addition to serum creatinine. Thus the inclusion of these other measurements may have negated the lack of change in serum creatinine observed in Section 6.1.1.

Overall, this work shows that GFR estimated using a creatinine-based equation should not be recommended for a patient who is under active oncological management, in particular platinum-based chemotherapy. Models which include additional biological or haematological variables may produce unbiased estimated values for these patients. However, more analysis is needed to prove or disprove this claim. Ideally, nuclear tracer measured GFR should be performed if an accurate measurement of GFR is required. When this is not possible, caution should be taken when interpreting any estimated GFR values for patients after receiving chemotherapy as any true loss of renal function may not be observed in estimated GFR.

6.2 Examining unstable creatinine using longitudinal data

In this section, preliminary analysis carried out for patients with longitudinal creatinine data is discussed. The patients examined here were previously used in Section 6.1.2, where the total number of measurements rather than the value of the measurements were considered. Here, we examine in more detail the value of these repeat creatinine measurements and

their impact on estimating GFR. The majority of the work that is outlined in this section is exploratory and to be developed in the future, where ultimately we aim to assess whether multiple creatinine measurements provide further information about a patient's GFR compared with single measurements.

In this exploratory analysis the performance of the refitted CamGFR model using longitudinal data is assessed. This is carried out, first by categorising patients as to whether they have increasing, decreasing or stable creatinine at the time of their nmGFR measurement. The performance of the refitted CamGFR model between these categorisations is then assessed. The model is then further assessed, using different input creatinine values based upon the timing of their measurement compared with the timing of the nmGFR measurement, for the different categorisations. These values are also compared against a previously published model, which adjusts creatinine based upon two subsequent creatinine measurements. Finally, possible future analyses using both these findings and our data is discussed.

6.2.1 Previous publications examining GFR when creatinine is unstable

Despite the availability of longitudinal creatinine data, estimating GFR using longitudinal data does not appear to have been extensively researched. Indeed, we only identified a few publications which explored the issues of estimating GFR when a patient's creatinine is changing between two subsequent measurements. In this section we examine two such publications, the first by Jelliffe et al. [64] who developed a model to estimate GFR using two creatinine measurements, the second by Chen [17] who suggested an adjustment based on two subsequent creatinine measurements for any creatinine-based equation. These papers were published in 1972 and 2013, respectively.

In addition to producing a model for creatinine clearance when a patient's creatinine is stable [62, 63], which was discussed in Section 2.4.3. Jelliffe et al. also produced a model that can be applied when a patient's creatinine is unstable [64]. This model requires two consecutive creatinine measurements, and uses the average and difference of these measurements along with the patient's sex, age and weight in a model for creatinine clearance. As is the case for the Jelliffe model for stable creatinine (see Section 2.4.3), this model is likely to be out of date given changes to the clinical methodology for measuring serum creatinine.

More recently, Chen [17] derived a theoretical equation, referred to as Kinetic eGFR (KeGFR), that adjusts the estimated GFR or creatinine clearance value based upon two subsequent creatinine measurements and steady state creatinine. This derivation was based on the relationship between the serum concentration of creatinine and its clearance rate in a dynamic system. Rather than directly estimating GFR, the equation provides an adjustment for an estimated GFR value produced by another model. Although this publication was well received with a number of citations in the literature, little validation could be found. The exception being Pianta et al. [99] and Dewitte et al. [31] which showed its advantage in predicting delayed graft function after kidney transplantation and predicting recovery from acute kidney injury respectively. The derived equation is shown below:

$$\text{KeGFR} = \text{eGFR}\frac{\text{SSCr}}{\text{SCr}_{\text{mean}}}\left(1 - \frac{\Delta\text{SCr}}{\Delta\text{T} \times \text{Max}\Delta\text{SCr}}\right) \tag{6.1}$$

where

- SSCr is the serum creatinine steady state value
- eGFR is the estimated GFR from a creatinine-based model
- SCr_{mean} is the mean of the two creatinine values
- ΔSCr is the change in the creatinine values
- ΔT is the time difference in days between the creatinine measurements
- MaxΔSCr is the theoretical maximum change in serum creatinine possible for the patients, which is typically set to be $1.5\,\text{mg}/\text{dL}$.

6.2.2 Exploratory analysis

For the purpose of this exploratory analysis, patients were examined who had at least one creatinine measurement in the 14 days leading up to, and after, their nmGFR measurement. Within our data, there were a total of 1,805 patients and 2,041 nmGFR measurements that met this criterion. First, patients were categorised as to whether their closest creatinine value before nmGFR was increasing, decreasing or not changing. For a patient to be categorised as having increasing creatinine we required their creatinine to increase by an average of 5% per day between the two closest measurements either side of their nmGFR (this is the same as log(SCr) increasing by log(1.05) = 0.049), and that the difference in the median value of up to three measurements either side of their nmGFR was increasing by at least 5%. For a

'decreasing' categorization, the same criteria were applied with obvious changes. Of the 2,041 nmGFR measurements in our data, 37 measurements were classified as increasing, 65 were classified as decreasing, and the remaining data were classified as not changing or stable.

Accuracy and bias of the refitted CamGFR model between increasing, decreasing and not changing creatinine was then examined. Table 6.2 provides the values of the four statistics used to compare models along with the median eGFR value and the number of patients in each category. The estimated GFR value for patients who were increasing was lower on average compared with patients with stable creatinine with the opposite true for patients with decreasing creatinine. We found that patients who had increasing creatinine were biased to have their GFR overestimated, with a median residual of -7.96 (95% CI -15.5 to -0.06). Patients with decreasing creatinine showed no significant bias. Overall, accuracy was far worse for both increasing and decreasing creatinine compared with not changing creatinine, with the RMSE (20.42 and 20.96 compared with 15.16 $\mathrm{mL/min}$) and P20 (0.459 and 0.4 compared with 0.255) significantly higher. This initial finding implies that additional caution should be taken when estimating GFR for a patient with increasing or decreasing creatinine. In the long-term one may consider incorporating the additional uncertainty of the estimated value into the model.

This comparison of the categorised creatinine values was then extended. For each categorisation, we compared the input of different creatinine values into refitted CamGFR depending on the timing of the creatinine measurement in relation to the nmGFR measurement. Overall, the input of three different creatinine values were compared: the closest measurement preceding nmGFR (CamGFR before), the closest measurement following (CamGFR after), and the measurement that was overall closest to (CamGFR closest). Where CamGFR closest is what we have used so far in this thesis. We also compared using these three inputs for refitted CamGFR to the KeGFR adjusted value, which adjusts for changing creatinine. For KeGFR, we used the median of measurements in the 60 day window around nmGFR (30 days before and 30 after) as the value for steady state creatinine. The statistics for each of these are provided in Table 6.1 and boxplots of the residuals are provided in Figure 6.5.

For patients who were categorised with stable creatinine, no difference was seen in the accuracy or bias between CamGFR before, CamGFR after and CamGFR closest. However, for patients with increasing creatinine the estimated values were lower for CamGFR after and higher for CamGFR before, which resulted in underestimate and overestimate of GFR respectively. The opposite was true for decreasing creatinine (Figure 6.5). These results were unsurprising, given that, by definition, patients with increasing creatinine must have a lower value of creatinine

equation	*n*	median eGFR	RMSE	median residual	IQR	P20
Decreasing						
CamGFR after	65	89.18	20.73 (17.19, 24.46)	-8.77 (-14.21, -4.08)	20.98 (12.45, 27.96)	0.462 (0.341, 0.583)
CamGFR closest	65	83.00	20.96 (17.40, 24.80)	1.07 (-2.71, 6.44)	23.75 (15.47, 31.32)	0.400 (0.282, 0.519)
CamGFR before	65	77.33	21.44 (18.01, 25.09)	4.47 (-0.95, 8.95)	22.48 (13.09, 29.84)	0.400 (0.280, 0.522)
CamGFR KeGFR	65	80.95	27.41 (22.77, 32.20)	0.70 (-5.02, 5.60)	31.58 (20.75, 43.60)	0.492 (0.369, 0.615)
Increasing						
CamGFR closest	37	76.08	20.42 (14.96, 26.49)	-7.96 (-15.46, -0.06)	22.02 (15.76, 31.68)	0.459 (0.301, 0.621)
CamGFR after	37	62.85	22.40 (16.47, 29.01)	6.11 (0.12, 14.86)	17.88 (-1.63, 29.87)	0.405 (0.252, 0.560)
CamGFR KeGFR	37	64.93	23.57 (17.33, 30.58)	4.12 (-5.67, 12.25)	28.23 (19.74, 40.65)	0.514 (0.359, 0.669)
CamGFR before	37	88.50	24.53 (18.93, 30.56)	-14.43 (-18.09, -10.67)	15.83 (4.68, 25.07)	0.541 (0.385, 0.699)
Not changing						
CamGFR after	1939	81.10	15.11 (14.48, 15.75)	-0.96 (-1.75, -0.24)	18.13 (17.04, 19.14)	0.259 (0.239, 0.278)
CamGFR closest	1939	80.76	15.16 (14.53, 15.81)	-0.66 (-1.49, 0.10)	17.78 (16.80, 18.87)	0.255 (0.236, 0.275)
CamGFR before	1939	80.96	15.29 (14.60, 15.99)	-0.14 (-0.82, 0.67)	17.56 (16.53, 18.57)	0.258 (0.239, 0.277)
CamGFR KeGFR	1939	78.70	18.00 (17.15, 18.87)	0.46 (-0.29, 1.32)	19.83 (18.83, 21.15)	0.317 (0.297, 0.338)

Table 6.2 Comparison of refitted CamGFR for three different creatinine values: closest measurement preceding (CamGFR before), the closest measurement following (CamGFR after), and the measurement that was overall closest to (CamGFR closest) the patient's nmGFR, along with the KeGFR adjusted estimated value (CamGFR KeGFR). Patients were compared in three separate groups, those with increasing, decreasing or stable creatinine. Details on the number of patients in each group (*n*), the median estimated GFR value (median eGFR), the root-mean-squared error (RMSE), the median residual, the residual interquartile range (IQR) and the proportion of samples with an absolute error greater than 20% (P20) are provided. 95% confidence intervals are provided in parentheses.

before the nmGFR compared with after. Further comparing the KeGFR adjusted CamGFR with the three different input values we found that although the estimated GFR is unbiased, accuracy of the estimation is decreased for the all three categorisations. The RMSE increases from 15.16, 20.42 and 20.96 to 18.00, 23.47 and 27.41 for stable, increasing and decreasing creatinine respectively.

Overall, this analysis has shown that when patients have increasing or decreasing creatinine the accuracy of the refitted CamGFR model is impaired. One may consider introducing additional uncertainty of the estimated value but otherwise care should be taken. We have also seen that using a KeGFR adjustment does lead to less biased estimated values for GFR, however this is at the expense of reduced accuracy, particularly for patients who do not have changing creatinine. This analysis has been performed using the refitted CamGFR model. This was chosen as it is unbiased for IDMS-traceable and non-IDMS-traceable creatinine, which were both present in this dataset. However, similar results were seen for other models which were overall unbiased in either IDMS-traceable or non-IDMS-traceable creatinine, in particular, CKD-EPI for non-IMDS creatinine and Lund-Malmo for IDMS-traceable creatinine.

Of course there are a number of considerations that we should take into account when evaluating these results. One of these considerations is to question the accuracy of the methodology used

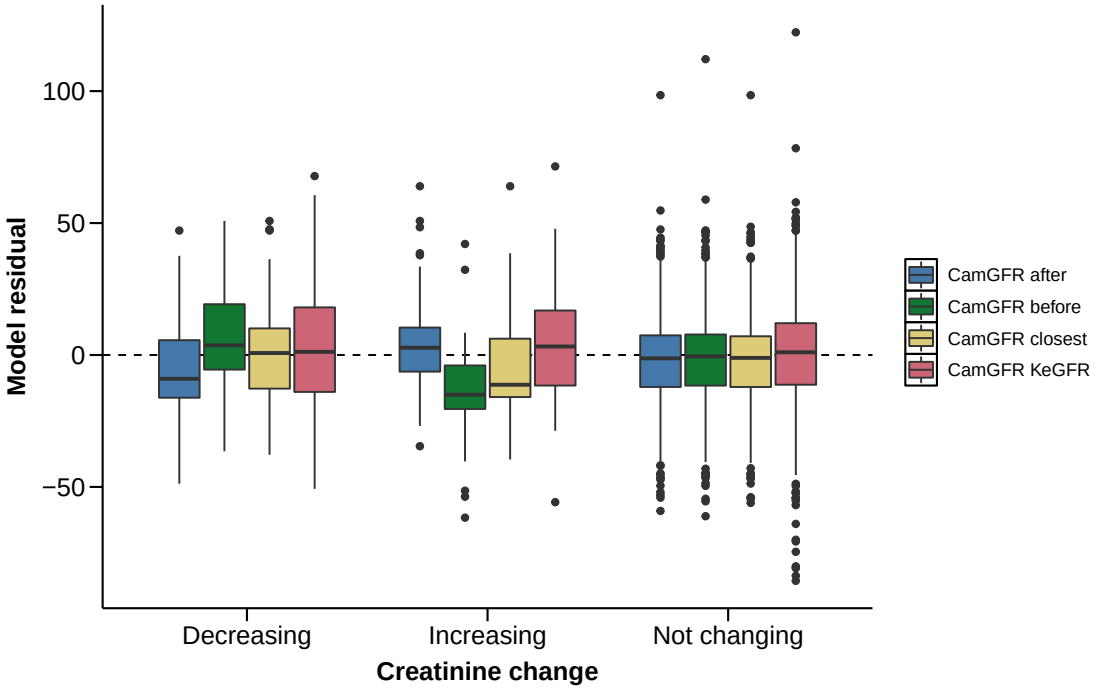


Figure 6.5 Boxplot of residuals for refitted CamGFR and the KeGFR adjusted CamGFR. Patients are grouped by whether there creatinine was categorised as increasing, decreasing or not changing at the time of the nmGFR measurements. Four different options are displayed for which input to use in the refitted CamGFR: the temporally closest measurements to the nmGFR, the temporally closest measurement which is prior to the nmGFR, the temporally closest measurement which is subsequent to the nmGFR, and the KeGFR adjusted to refitted CamGFR.

to categorise patients as having increasing, decreasing or stable creatinine. We reviewed the plots for individual patients who were categorised as having increasing GFR, Figure 6.6. In these plots some clear evidence of increasing creatinine is observed, whereby the plots exhibit several low values for creatinine which then steadily increase during the time period when the nmGFR was measured. Examples of such plots are seen for patients 1, 2, 6, 8 and 13. Conversely, for some plots the categorisation of increasing creatinine is less clear-cut. One example of this is observed in the plot for patient 34, where the creatinine values vary during the time-period without a consistent trend. Further examples where the trend for creatinine is less clear-cut can be seen in plots for patients 29, 30, 32 and 35. Additionally, comparing the residuals of the patients who show a clearer increase in creatinine to those who do not, we see that those with a clearer increase tend to be overestimated (large negative residuals). Thus the overestimate of refitted CamGFR when a patient has increasing creatinine may be more significant if patients were more accurately categorised.

The plot, and associated results, for patient 37 also raise an interesting consideration. The patient's creatinine increases from a value of 0.5 mg/dL, eight days before their nmGFR, to 2 mg/dL, 8 days after. The creatinine value after nmGFR was measured in closest time proximity, where the value of this patient's nmGFR was 90 mL/dL. The residual shown in the plot corresponds to using the creatinine value after nmGFR, which estimated GFR to be 26 mL/dL. This is in contrast to the result observed when using the creatinine value measure prior to nmGFR with refitted CamGFR, which produced an estimated value for GFR of 93.2 mL/dL. What is likely to have occurred is that the patient's GFR decreased (for reason such as an acute kidney injury) in the eight days after their nmGFR, causing their creatinine to increase. Unfortunately our data does not have creatinine measurements for these eight days that would confirm this hypothesis.

Another consideration is the method used to calculate the steady state of each patient's serum creatinine. As described above, for this calculation the median creatinine value measured in the 30 days before and after nmGFR was used. However, this may be an inaccurate estimation for many of the patients. Again examining the plots in Figure 6.6, the line for the median value can be some distance from where the steady state for some individual patients would be observed. Patient 6, for example, appears to have a steady state creatinine value near their minimum observed value, however the median value line is positioned for a higher value than this. Moreover, for some other patients who have few serum creatinine measurements, there is not enough data to be able to estimate what their steady state creatinine value would be and so again the median value could be some distance for their steady state creatinine.

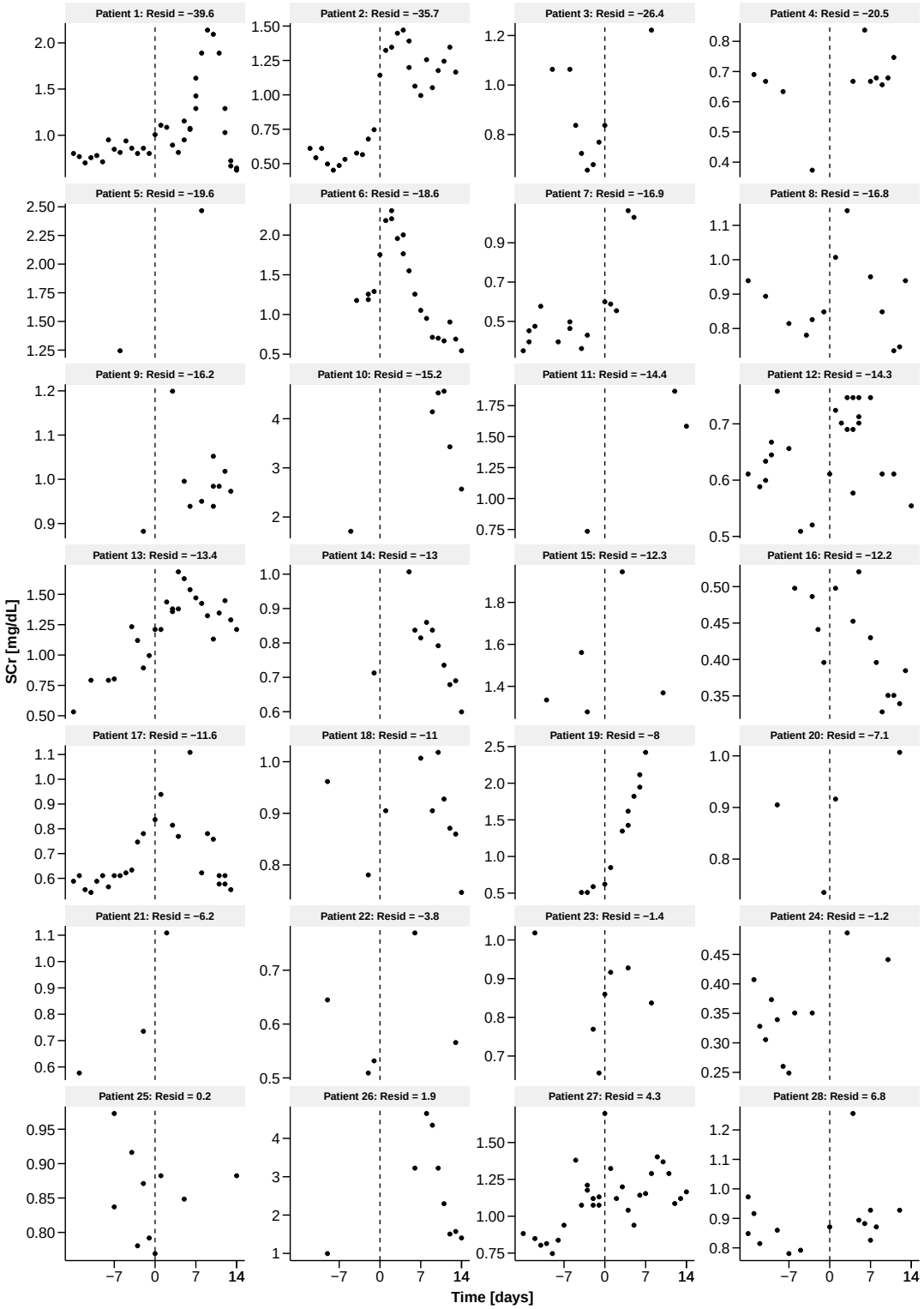


Figure 6.6 Serum creatinine time series of individuals who were categorised as having increasing serum creatinine at the time of their nmGFR measurement. All creatinine values measured within 14 days either side of the nmGFR are included. Patients are ordered by the residual form the refitted CamGFR model (displayed above each time series) when using the closest creatinine measurement to the nmGFR.

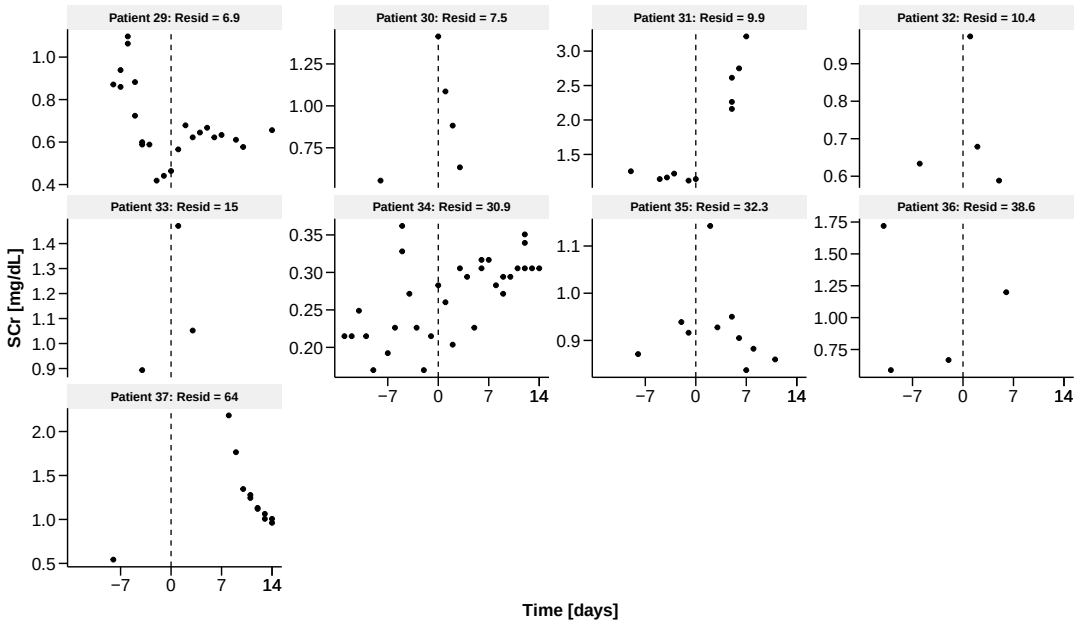


Figure 6.6 Serum creatinine time series of individuals who were categorised as having increasing serum creatinine at the time of their nmGFR measurement. All creatinine values measured within 14 days either side of the nmGFR are included. Patients are ordered by the residual form the refitted CamGFR model (displayed above each time series) when using the closest creatinine measurement to the nmGFR.

6.2.3 Direction of future work

This section has introduced exploratory analysis of longitudinal creatinine data, which we intend to develop in the future to examine whether multiple creatinine measurements provide further information about a patient's GFR compared with single measurements. One way in which this may be done is to explore the incorporation of several creatinine values directly in the refitted CamGFR model to better estimate GFR and the variability of this estimated value. If a patient's creatinine is stable but fluctuating then additional measurements of creatinine may be used to model the variably of the measurement which could then be translated to the variability in the estimated GFR. We may then also explore the incorporation of these additional creatinine values alongside the additional biochemical and haematological blood tests that were investigated in Chapter 5, as this may further improve the accuracy and bias of the model. This work may also by further extended through examining its application to liver function where we would model the relationship between the kidney and the liver using longitudinal measurements that are associated with both kidney and liver function. To carry this out, we would ideally like patient data that are regularly and routinely collected with multiple nmGFR

measurements. This would provide some mitigation of the considerations we have raised in this section. However, data collation of this nature is not often the case in clinical medicine and so this would need to be taken into account in any future work.

6.3 Comparison of measured GFR between patients with and without cancer

A final question that has arisen through the work in the thesis is addressed in this section: do patients with cancer have normal GFR for their age compared with patients who are healthy? In Section 4.1, we alluded to the fact that the average GFR values in our dataset are higher than in other datasets used to develop previous models. This was often the case as other datasets were enriched with chronic kidney disease patients who would typically have values for GFR that are lower than would be expected for healthy patients of their age. It is therefore unclear from these data whether patients with cancer have the same normal GFR as healthy patients of the same age. It is challenging to acquire the data necessary to examine this question because healthy patients do not usually have their nmGFR measured unless they are candidates to donate a kidney. In our dataset, we had a small subset of such patients (n = 188), though to compare these patients with patients with cancer (n = 7,849) we needed more data. Efforts were therefore made to supplement this data consisting of seemingly healthy patients who are candidates for kidney transplant surgery. The relationship between GFR and age for patients with and without cancer was then compared to examine whether patients with cancer have normal GFR for their age. This was particularly important to assess when considering use of the refitted CamGFR model to estimate GFR for patients who are healthy.

The supplementary data were acquired from Fenton et al. [38] whose publication focused on assessing the current GFR threshold values for kidney donations. Multiple attempts were made to contact the author and acquire data directly, however this proved unsuccessful. The relevant data (GFR and age) from Figure 1 in the publication was therefore manually extracted. This led us to acquire additional data from 3,105 prospective liver kidney donors. There were actually data from 2,974 patients in Fenton et al. however the resolution of Figure 1 led to some duplication when the plot was extracted. The resulting final dataset consisted of 7,849 and 3,293 patients with and without cancer respectively.

Using these additional data, the relationship between GFR and age for patients with and without cancer was compared. Both of these patient types showed a relatively flat relationship between GFR and age for the age range of 18 to 40 years, with a steady decline in GFR after the age of 40. Using the **segmented** R package a piecewise linear model was fitted for all of the data to estimate where the break point was. The estimated break point was calculated as 40.1 years. With this break point separate piecewise linear models were then fitted for the patients with and without cancer, which allowed for an estimation of the difference in slope between these two patient types both before and after the break point.

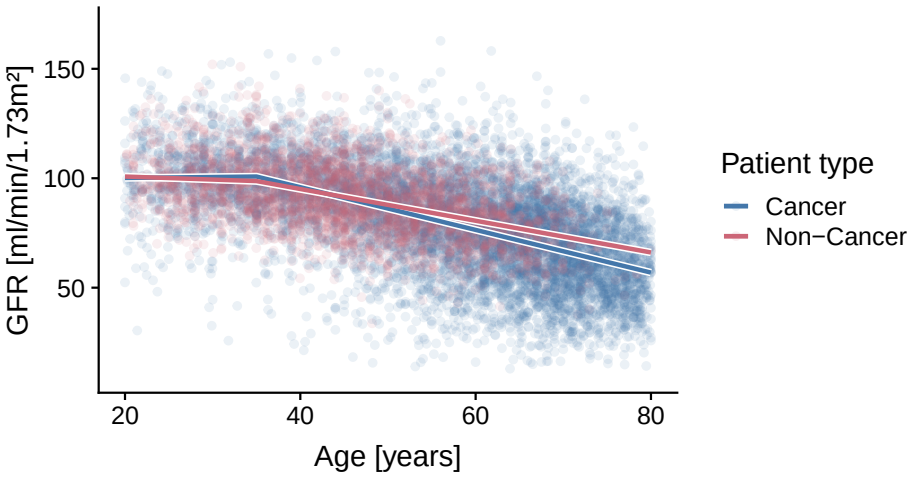


Figure 6.7 GFR against age for patients with and without cancer. The majority of the non-cancer patients were extracted from [38, Figure 1]. The fitted curves are piecewise linear with a break point at 40.5 years.

The difference in the slope of the relationship before the break point was estimated to be 0.048 (cancer: -0.326; non-cancer: -0.278; p-value: 0.056), whereas the difference after the break point was 0.253 (cancer: -1.014; non-cancer: -0.761; p-value: <0.0001). The plot of GFR against age for patients with and without cancer, along with these fitted piecewise linear relationships are shown in Figure 6.7. It appears that for patients under the age of 40, there is little difference in GFR between patients with and without cancer, whilst after the age of 40 the difference in GFR increases with age. For patients aged over 70 the mean GFR was 60.8 $\mathrm{mL/min/1.73m^2}$ for patients with cancer compared with 70.0 $\mathrm{mL/min/1.73m^2}$ for patients without cancer (t.test p-value < 0.0001).

For the majority of the age spectrum the differences in GFR between patients with cancer and healthy patients are minor, however, as age progresses the disparity between measured GFR is increasingly noticeable. Another way to interpret this finding is that patients with cancer, even

early in their diagnosis as the majority of the patients in our dataset are, have a kidney function that corresponds to an older person who does not have cancer. An associative study like this does not allow us to distinguish as to whether ageing and cancer cause loss of renal function by the same or different mechanisms, or even the direction of causality. For example, patients with impaired renal function may be more likely to develop cancer or there may be a latent variable which predisposes both an increased likelihood of developing cancer and a decline in renal function. Another possibility is that the observed difference is driven by the small number of patients who may have already received chemotherapy prior to their nmGFR, which was observed in Section 6.1, to be associated with declined renal function. Treatment-annotated data needs to be obtained to rule out this possibility. On the basis of these results we conclude that there is good evidence for the applicability of refitted CamGFR for patients up to the age of 60, where after this age more caution should be taken.

Chapter 7

Conclusion

Many aspects of the management of patients with cancer are informed by knowledge of GFR. The purpose of this thesis was to review previously published models for estimating GFR and, more importantly, to develop a new regression-based model to best estimate GFR for patients with cancer. Our work was then further developed, first by considering the inclusion of additional biochemical and haematological data in a regression-based model for GFR, and second by critically assessing a model against clinical considerations that should be taken into account for any estimation of a patient's GFR. We had a clear rationale for both reviewing and building upon existing models. Most notably, we identified limited large-scale studies of existing models and their applicability to patients with cancer. In addition, we identified the need to consider the increasing availability and volume of routinely measured patient data, alongside changing methods to measure clinical variables.

We initially developed a model using data from a single centre to estimate GFR for patients with cancer, the CamGFR model. CamGFR models GFR against a patient's blood serum creatinine concentration as well as other demographic data. We showed, both in an internal and in a small scale external validation, that this model is more accurate than previously published models. Publication of this initial model enabled us to collate additional datasets from different cancer centres in order to further develop CamGFR.

Using this large multicentre dataset, we compared different methods to measure serum creatinine in order to evaluate whether different methods required adjustments to CamGFR. Overall, we found a significant difference in serum creatinine values dependent on whether they were IDMS-traceable or not. Given this result, and the fact that the initial CamGFR model was

developed in a single centre using only non-IDMS-traceable creatinine, we then sought to redevelop the model to be less centre dependant and suitable for both methodologies. The refitted CamGFR model expanded the initial model by including an interaction term indicating the creatinine method. The refitted CamGFR model was shown to be more accurate than previously published models for both IDMS-traceable and non-IDMS-traceable creatinine. Previously published models were observed to be biased in estimating GFR for one of the two methods. In particular, the CKD-EPI model (which was developed using IDMS-traceable creatinine) was highly biased for IDMS-traceable creatinine and unbiased for non-IDMS-traceable creatinine. We have worked to make this model accessible and usable in day-to-day clinical practice by developing an online application that enables users to input patient data and receive not only an estimate of the patient's GFR but also the user defined confidence interval for this prediction.

Of course, the refitted CamGFR model was not without limitations, which were mainly concerned with the population demographics and data availability. Serum creatinine was used as the filtration marker in CamGFR, however, some recent literature has been focused on the use of cystatin C in conjunction with, or instead of, serum creatinine. Cystatin C was not measured in any of the cancer centres when the data were collected, and so was not available for the purpose of this thesis. Our work was also limited by a lack of full annotation with regards to treatment history and also by a lack of ethnic diversity, as data were predominately from white Caucasian patients. Future publications may encourage further collaborations and data sharing so that we may address these limitations and further validate or adjust the model.

We continued our analysis of GFR in patients with cancer by examining whether including other routine biochemical and hematological variables in the model would further improve the estimation of GFR. In doing so we found that some variables were consistently observed to be predictive of GFR, in particular urea, albumin, ALP and ALT. In addition, variables associated with red blood cells were also predictive of GFR. Including these variables improved estimations of GFR compared with creatinine only models in both an internal and small external validation. These fitted relationships further provided interesting observations about the underlying biology and the correlation between different organ systems, in particular the liver and kidney. The main limitation for this exploratory study was, once again, that it relied upon data from a single cancer centre (*Cambridge*) along with a small number of patients from a secondary centre. This limitation is particularly important to take into account as we suspected that many of these routine measurements were likely to have centre dependant biases. Thus to further develop this study, we will seek to obtain from other centres the same data

as we had available from *Cambridge*. If, using additional multicentre data, further validation corroborates the current finding, we will then aim to develop a final model.

Over the course of our work, through examining literature, conducting analyses, and speaking with medical and statistical professionals, we came across a number of important clinical considerations for estimating GFR in patients with cancer. We first considered the effect of chemotherapy and other treatment on GFR, creatinine and estimated GFR. A key challenge for this analysis was a lack of data annotation providing detail on a patient's treatment history. To resolve this we used two surrogate methods to distinguish patients who had likely received chemotherapy treatment versus those who had not. Similar to previous studies we found that after receiving treatment a patient's GFR was reduced, which was not initially matched by an increase in serum creatinine. This resulted in an overestimate of GFR for patients under active oncological management, such as currently receiving, or post chemotherapy. We corroborated this finding by examining patients who had received numerous serum creatinine blood tests in the two weeks preceding their nmGFR; these patients also had their GFR overestimated by creatinine-based models.

Another consideration was the use of creatinine-based models for patients who had unstable creatinine values. Using longitudinal serum creatinine measurements we categorised patients as to whether their creatinine was increasing, decreasing or stable at the time of their nmGFR. We found that patients with increasing creatinine had an overestimated GFR, whereas no bias was observed for estimated GFR when a patient's creatinine was decreasing. The limiting factor with this analysis was the accuracy of the categorisation of patients, which was again compounded by a lack of consistency in the data available, a factor over which we had no control.

A final consideration was to assess the applicability of using CamGFR to estimate GFR for healthy patients. We compared GFR in patients with cancer to healthy kidney donor patients and observed that for younger patients there was no difference in GFR, however the rate of decline in GFR with age was seen to be steeper for patients with cancer.

Even though this thesis presents the largest and most extensive dataset collected and analysed with regards to GFR for patients with cancer, a recurring limiting factor has been a lack of data availability and data consistency. This is not because such data does not exist, indeed patient data are regularly used for clinical publications and is routinely collected in hospitals or treatment centres. This field of research in general would benefit from increasing collaboration

particularly with regards to the sharing of data. This should be an international effort as the models produced from work such as that carried out in this thesis may be applied to patient diagnosis and treatment internationally. Care, of course, needs to be taken to ensure ethical standards are upheld, with patient anonymity and privacy respected. With the increasing volume, storage and usage of routine patient data we are entering an exciting time in research. In the near future, we should be able to use more data driven models to further our understanding of underlying biology and improve both patient care and outcomes.

To conclude, we have shown that it is feasible to collate retrospective clinical data from multiple centres which is of sufficient quality to conduct robust statistical modelling. This thesis is the first multicentre comparison of models to estimate GFR in patients with cancer. In developing a new statistical model, we address gaps in clinical research that have emerged due to increasing data availability, the evolution of methods used to measure clinical variables such as creatinine, and limited research into models that are developed based on patients with cancer. Overall, we have demonstrated that CamGFR improves the accuracy of existing models for patients with and without cancer, in turn, this improved accuracy may contribute to improved patient care.

References

[1] Abdelhafiz, A. H., Brown, S. H., Bello, A., and El Nahas, M. (2010). Chronic kidney disease in older people: Physiology, pathology or both? *Nephron Clinical Practice*, 116:19–24.

[2] Ainsworth, N. L., Marshall, A., Hatcher, H., Whitehead, L., Whitfield, G. A., and Earl, H. M. (2012). Evaluation of glomerular filtration rate estimation by Cockcroft-Gault, Jelliffe, Wright and Modification of Diet in Renal Disease (MDRD) formulae in oncology patients. *Annals of Oncology*, 23:1845–1853.

[3] Anastasio, P., Cirillo, M., Spitali, L., Frangiosa, A., Pollastro, R. M., and De Santo, N. G. (2001). Level of hydration and renal function in healthy humans. *Kidney International*, 60:748–756.

[4] Barai, S., Bandopadhayaya, G. P., Patel, C. D., Rathi, M., Kumar, R., Bhowmik, D., Gambhir, S., Gopendro Singh, N., Malhotra, A., and Gupta, K. D. (2005). Do healthy potential kidney donors in India have an average glomerular filtration rate of 81.4 ml/min? *Nephron Physiology*, 101:21–26.

[5] Barraclough, L., Field, C., Wieringa, G., Swindell, R., Livsey, J., and Davidson, S. (2008). Estimation of renal function — what is appropriate in cancer patients? *Clinical Oncology*, 20:721–726.

[6] Biörk, J., Carlson, J., Bäck, S., Simonsson, P., Bakoush, O., Grubb, A., Lindström, V., Nyman, U., and Sterner, G. (2007). Prediction of relative glomerular filtration rate in adults: New improved equations based on Swedish Caucasians and standardized plasma-creatinine assays. *Scandinavian Journal of Clinical and Laboratory Investigation*, 67:678–695.

[7] Björk, J., Grubb, A., Sterner, G., and Nyman, U. (2011). Revised equations for estimating glomerular filtration rate based on the Lund-Malmö Study cohort. *Scandinavian Journal of Clinical and Laboratory Investigation*, 71:232–239.

[8] Bökenkamp, A., Laarman, C. A. R. C., Braam, K. I., van Wijk, J. A. E., Kors, W. A., Kool, M., de Valk, J., Bouman, A. A., Spreeuwenberg, M. D., and Stoffel-Wagner, B. (2007). Effect of corticosteroid therapy on low-molecular weight protein markers of kidney function. *Clinical Chemistry*, 53:2219–2221.

[9] Boutten, A., Bargnoux, A.-S., Carlier, M.-C., Delanaye, P., Rozet, E., Delatour, V., Cavalier, E., Hanser, A.-M., Froissart, M., Cristol, J.-P., and Piéroni, L. (2013). Enzymatic but not compensated Jaffe methods reach the desirable specifications of NKDEP at normal levels of

creatinine. Results of the French multicentric evaluation. *Clinica Chimica Acta*, 419:132–135.

[10] Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26:211–252.

[11] British Transplantation Society (2018). Guidelines for Living Donor Kidney Transplantation. Technical report, British Transplantation Society.

[12] Burman, P. (1989). A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika*, 76:503–514.

[13] Calvert, A. H., Newell, D. R., Gumbrell, L. A., Reilly, S. O., Burnell, M., Boxall, F. E., Siddik, Z. H., Judson, I. R., Gore, M. E., and Wiltshaw, E. (1989). Carboplatin dosage: prospective evaluation of a simple formula based on renal function. *Journal of Clinical Oncology*, 7:1748–1756.

[14] Cancer Research UK (2019). Cancer incidence by age.

[15] Cathomas, R., Klingbiel, D., Geldart, T. R., Mead, G. M., Ellis, S., Wheater, M., Simmonds, P., Nagaraj, N., von Moos, R., and Fehr, M. (2014). Relevant risk of carboplatin underdosing in cancer patients with normal renal function using estimated GFR: lessons from a stage I seminoma cohort. *Annals of Oncology*, 25:1591–1597.

[16] Chang, W., Cheng, J., Allaire, J., Xie, Y., and McPherson, J. (2019). *shiny: Web Application Framework for R*. R package version 1.3.2.

[17] Chen, S. (2013). Retooling the creatinine clearance equation to estimate kinetic gfr when the plasma creatinine is changing acutely. *Journal of the American Society of Nephrology*, 24:877–88.

[18] Cheung, K. L. and Lafayette, R. A. (2013). Renal physiology of pregnancy. *Advances in Chronic Kidney Disease*, 20:209–214.

[19] Chudleigh, R. A., Ollerton, R. L., Dunseath, G., Peter, R., Harvey, J. N., Luzio, S., and Owens, D. R. (2008). Performance of the revised '175' Modification of Diet in Renal Disease equation in patients with type 2 diabetes. *Diabetologia*, 51:1714–1718.

[20] Cimerman, N., Brguljan, P. M., Krašovec, M., Šuškovič, S., and Kos, J. (2000). Serum cystatin C, a potent inhibitor of cysteine proteinases, is elevated in asthmatic patients. *Clinica Chimica Acta*, 300:83–95.

[21] Clogg, C. C., Petkova, E., and Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American Journal of Sociology*, 100:1261–1293.

[22] Cockcroft, D. W. and Gault, M. H. (1976). Prediction of creatinine clearance from serum creatinine. *Nephron*, 16:31–41.

[23] Coresh, J., Astor, B. C., McQuillan, G., Kusek, J., Greene, T., Van Lente, F., and Levey, A. S. (2002). Calibration and random variation of the serum creatinine assay as critical elements of using equations to estimate glomerular filtration rate. *American Journal of Kidney Diseases*, 39:920–929.

[24] Craig, A. J., Samol, J., Heenan, S. D., Irwin, A. G., and Britten, A. (2012). Overestimation of carboplatin doses is avoided by radionuclide GFR measurement. *British Journal of Cancer*, 107:1310–1316.

[25] Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press.

[26] De Boer, I. H. and Utzschneider, K. M. (2017). The kidney's role in systemic metabolism - Still much to learn. *Nephrology Dialysis Transplantation*, 32:588–590.

[27] De Santis, M., Bellmunt, J., Mead, G., Kerst, J. M., Leahy, M., Maroto, P., Gil, T., Marreaud, S., Daugaard, G., Skoneczna, I., Collette, S., Lorent, J., De Wit, R., and Sylvester, R. (2011). Randomized phase II/III trial assessing gemcitabine/carboplatin and methotrexate/carboplatin/vinblastine in patients with advanced urothelial cancer who are unfit for cisplatin-based chemotherapy: EORTC study 30986. *Journal of Clinical Oncology*, 30:191–199.

[28] Delanaye, P., Cavalier, E., and Pottel, H. (2017). Serum creatinine: Not so simple! *Nephron*, 136:302–308.

[29] Delanaye, P., Pieroni, L., Abshoff, C., Lutteri, L., Chapelle, J.-P., Krzesinski, J.-M., Hainque, B., and Cavalier, E. (2008). Analytical study of three cystatin C assays and their impact on cystatin C-based GFR-prediction equations. *Clinica Chimica Acta*, 398:118–124.

[30] Delanaye, P., Schaeffner, E., Ebert, N., Cavalier, E., Mariat, C., Krzesinski, J.-M., and Moranne, O. (2012). Normal reference values for glomerular filtration rate: what do we really know? *Nephrology Dialysis Transplantation*, 27:2664–2672.

[31] Dewitte, A., Joannès-Boyau, O., Sidobre, C., Fleureau, C., Bats, M. L., Derache, P., Leuillet, S., Ripoche, J., Combe, C., and Ouattara, A. (2015). Kinetic eGFR and novel AKI biomarkers to predict renal recovery. *Clinical Journal of the American Society of Nephrology*, 10:1900–1910.

[32] du Bois, A., Herrstedt, J., Hardy-Bessard, A.-C., Müller, H.-H., Harter, P., Kristensen, G., Joly, F., Huober, J., Avall-Lundqvist, E., Weber, B., Kurzeder, C., Jelic, S., Pujade-Lauraine, E., Burges, A., Pfisterer, J., Gropp, M., Staehle, A., Wimberger, P., Jackisch, C., and Sehouli, J. (2010). Phase III trial of carboplatin plus paclitaxel with or without gemcitabine in first-line treatment of epithelial ovarian cancer. *Journal of Clinical Oncology*, 28:4162–4169.

[33] Du Bois, D. and Du Bois, E. F. (1916). A formula to estimate the approximate surface area if height and weight be known. *Archives of Internal Medicine*, XVII:863–871.

[34] Edwards, K. D. and Whyte, H. M. (1959). Plasma creatinine level and creatinine clearance as tests of renal function. *Australasian Annals of Medicine*, 8:218–224.

[35] Edwards, M. S., Wilson, D. B., Craven, T. E., Stafford, J., Fried, L. F., Wong, T. Y., Klein, R., Burke, G. L., and Hansen, K. J. (2005). Associations between retinal microvascular abnormalities and declining renal function in the elderly population: The cardiovascular health study. *American Journal of Kidney Diseases*, 46:214–224.

[36] Efron, B. and Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. Chapman & Hall, New York.

[37] Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11:89–121.

[38] Fenton, A., Montgomery, E., Nightingale, P., Peters, A. M., Sheerin, N., Wroe, A. C., and Lipkin, G. W. (2018). Glomerular filtration rate: new age- and gender-specific reference ranges and thresholds for living kidney donation. *BMC Nephrology*, 19:336.

[39] Fricker, M., Wiesli, P., Brändle, M., Schwegler, B., and Schmid, C. (2003). Impact of thyroid dysfunction on serum cystatin C. *Kidney International*, 63:1944–1947.

[40] Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33:1–22.

[41] Gansevoort, R. T., Correa-Rotter, R., Hemmelgarn, B. R., Jafar, T. H., Heerspink, H. J. L., Mann, J. F., Matsushita, K., and Wen, C. P. (2013). Chronic kidney disease and cardiovascular risk: epidemiology, mechanisms, and prevention. *The Lancet*, 382:339–352.

[42] Geddes, C. C., Woo, Y. M., and Brady, S. (2008). Glomerular filtration rate - What is the rationale and justification of normalizing GFR for body surface area? *Nephrology Dialysis Transplantation*, 23:4–6.

[43] Giglio, D. (2014). A new equation for estimating glomerular filtration rate in cancer patients. *Chemotherapy*, 60:63–72.

[44] Goldie, D. J. and McConnell, A. A. (1990). Serum alanine transaminase (ALT) reference ranges estimated from blood donors. *Journal of Clinical Pathology*, 43:929–931.

[45] Gowda, S., Desai, P. B., Hull, V. V., Math, A. A. K., Vernekar, S. N., and Kulkarni, S. S. (2009). A review on laboratory liver function tests. *The Pan African Medical Journal*, 3:17.

[46] Grubb, A. (2017). Cystatin C is indispensable for evaluation of kidney disease. *EJIFCC*, 28:268–276.

[47] Grubb, A., Blirup-Jensen, S., Lindström, V., Schmidt, C., Althaus, H., Zegers, I., and IFCC Working Group on Standardisation of Cystatin C (WG-SCC). (2010). First certified reference material for cystatin C in human serum ERM-DA471/IFCC. *Clinical Chemistry and Laboratory Medicine*, 48:1619–1621.

[48] Grubb, A., Horio, M., Hansson, L. O., Björk, J., Nyman, U., Flodin, M., Larsson, A., Bökenkamp, A., Yasuda, Y., Blufpand, H., Lindström, V., Zegers, I., Althaus, H., Blirup-Jensen, S., Itoh, Y., Sjöström, P., Nordin, G., Christensson, A., Klima, H., Sunde, K., Hjort-Christensen, P., Armbruster, D., and Ferrero, C. (2014). Generation of a new cystatin C-based estimating equation for glomerular filtration rate by use of 7 assays standardized to the international calibrator. *Clinical Chemistry*, 60:974–986.

[49] Haase, M., Story, D. A., and Haase-Fielitz, A. (2011). Renal injury in the elderly: Diagnosis, biomarkers and prevention. *Best Practice and Research: Clinical Anaesthesiology*, 25:401–412.

[50] Hall, J. E. (2011). *Guyton and Hall: Textbook of Medical Physiology*. Saunders Elsevier, Philadelphia, 12th edition.

[51] Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer Series in Statistics, second edition.

[52] Haycock, G. B., Schwartz, G. J., and Wisotsky, D. H. (1978). Geometric method for measuring body surface area: A height-weight formula validated in infants, children, and adults. *The Journal of Pediatrics*, 93:62–66.

[53] Himmelfarb, J. and Sayegh, M. (2010). *Chronic Kidney Disease, Dialysis, and Transplantation*. Saunders Elsevier, third edition.

[54] Horio, M., Imai, E., Yasuda, Y., Watanabe, T., and Matsuo, S. (2013). GFR estimation using standardized serum cystatin C in Japan. *American Journal of Kidney Diseases*, 61:197–203.

[55] Hoste, L., Deiteren, K., Pottel, H., Callewaert, N., and Martens, F. (2015). Routine serum creatinine measurements: how well do we perform? *BMC Nephrology*, 16:21.

[56] Hughes, A., Calvert, P., Azzabi, A., Plummer, R., Johnson, R., Rusthoven, J., Griffin, M., Fishwick, K., Boddy, A. V., Verrill, M., and Calvert, H. (2002). Phase I clinical and pharmacokinetic study of pemetrexed and carboplatin in patients with malignant pleural mesothelioma. *Journal of Clinical Oncology*, 20:3533–3544.

[57] Husdan, H. and Rapoport, A. (1968). Estimation of creatinine by the Jaffe reaction: a comparison of three methods. *Clinical chemistry*, 14:222–238.

[58] Inker, L. A., Schmid, C. H., Tighiouart, H., Eckfeldt, J. H., Feldman, H. I., Greene, T., Kusek, J. W., Manzi, J., Van Lente, F., Zhang, Y. L., Coresh, J., and Levey, A. S. (2012). Estimating glomerular filtration rate from serum creatinine and Cystatin C. *New England Journal of Medicine*, 367:20–29.

[59] Jafar, T. H., Islam, M., Jessani, S., Bux, R., Inker, L. A., Mariat, C., and Levey, A. S. (2011). Level and determinants of kidney function in a South Asian population in Pakistan. *American Journal of Kidney Diseases*, 58:764–772.

[60] Janowitz, T., Williams, E. H., Marshall, A., Ainsworth, N., Thomas, P. B., Sammut, S. J., Shepherd, S., White, J., Mark, P. B., Lynch, A. G., Jodrell, D. I., Tavaré, S., and Earl, H. (2017). New model for estimating glomerular filtration rate in patients with cancer. *Journal of Clinical Oncology*, 35:2798–2805.

[61] Jelkmann, W. (2011). Regulation of erythropoietin production. *The Journal of Physiology*, 589:1251–1258.

[62] Jelliffe, R. W. (1971). Estimation of creatinine clearance when urine cannot be collected. *The Lancet*, 1:975–976.

[63] Jelliffe, R. W. (1973). Creatinine clearance: bedside estimate. *Annals of Internal Medicine*, 79:604–605.

[64] Jelliffe, R. W. and Jelliffe, S. M. (1972). A computer program for estimation of creatinine clearance from unstable serum creatinine levels, age, sex, and weight. *Mathematical Biosciences*, 14:17–24.

[65] Jodrell, D. I., Egorin, M. J., Canetta, R. M., Langenberg, P., Goldbloom, E. P., Burroughs, J. N., Goodlow, J. L., Tan, S., and Wiltshaw, E. (1992). Relationships between carboplatin exposure and tumor response and toxicity in patients with ovarian cancer. *Journal of Clinical Oncology*, 10:520–528.

[66] Johnson, D. H., Fehrenbacher, L., Novotny, W. F., Herbst, R. S., Nemunaitis, J. J., Jablons, D. M., Langer, C. J., DeVore, R. F., Gaudreault, J., Damico, L. A., Holmgren, E., and Kabbinavar, F. (2004). Randomized phase II trial comparing bevacizumab plus carboplatin and paclitaxel with carboplatin and paclitaxel alone in previously untreated locally advanced or metastatic non-small-cell lung cancer. *Journal of Clinical Oncology*, 22:2184–2191.

[67] Katsumata, N., Yasuda, M., Takahashi, F., Isonishi, S., Jobo, T., Aoki, D., Tsuda, H., Sugiyama, T., Kodama, S., Kimura, E., Ochiai, K., Noda, K., and Japanese Gynecologic Oncology Group (2009). Dose-dense paclitaxel once a week in combination with carboplatin every 3 weeks for advanced ovarian cancer: a phase 3, open-label, randomised controlled trial. *The Lancet*, 374:1331–1338.

[68] Keener, J. and Sneyd, J. (2009). *Mathematical Physiology II: Systems Physiology*. Springer, second edition.

[69] Koopman, M. G., Koomen, G. C., Krediet, R. T., de Moor, E. A., Hoek, F. J., and Arisz, L. (1989). Circadian rhythm of glomerular filtration rate in normal individuals. *Clinical Science*, 77:105–11.

[70] Kos, J., Stabuc, B., Cimerman, N., and Brünner, N. (1998). Serum cystatin C, a new marker of glomerular filtration rate, is increased during malignant progression. *Clinical Chemistry*, 44:2556–2557.

[71] Kunde, S. S., Lazenby, A. J., Clements, R. H., and Abrams, G. A. (2005). Spectrum of NAFLD and diagnostic implications of the proposed new normal range for serum ALT in obese women. *Hepatology*, 42:650–656.

[72] Laterza, O. F., Price, C. P., and Scott, M. G. (2002). Cystatin C: An improved estimator of glomerular filtration rate? *Clinical Chemistry*, 48:699–707.

[73] Lauritsen, J., Gundgaard, M. G., Mortensen, M. S., Oturai, P. S., Feldt-Rasmussen, B., and Daugaard, G. (2014). Reliability of estimated glomerular filtration rate in patients treated with platinum containing therapy. *International Journal of Cancer*, 135:1733–1739.

[74] Levey, A. S., Bosch, J. P., Lewis, J. B., Greene, T., Rogers, N., and Roth, D. (1999). A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. *Annals of Internal Medicine*, 130:461–470.

[75] Levey, A. S., Coresh, J., Greene, T., Marsh, J., Stevens, L. A., Kusek, J. W., and Van Lente, F. (2007). Expressing the modification of diet in renal disease study equation for estimating glomerular filtration rate with standardized serum creatinine values. *Clinical Chemistry*, 53:766–772.

[76] Levey, A. S., Coresh, J., Greene, T., Stevens, L. A., Zhang, Y. L., Hendriksen, S., Kusek, J. W., Van Lente, F., and Chronic Kidney Disease Epidemiology Collaboration. (2006). Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. *Annals of Internal Medicine*, 145:247–254.

[77] Levey, A. S., Stevens, L. A., Schmid, C. H., Zhang, Y. L., Castro, A. F., Feldman, H. I., Kusek, J. W., Eggers, P., Van Lente, F., Greene, T., Coresh, J., and CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration) (2009). A new equation to estimate glomerular filtration rate. *Annals of Internal Medicine*, 150:604–612.

[78] Lewis, J., Agodoa, L., Cheek, D., Greene, T., Middleton, J., O'Connor, D., Ojo, A., Phillips, R., Sika, M., and Wright, J. (2001). Comparison of cross-sectional renal function measurements in African Americans with hypertensive nephrosclerosis and of primary formulas to estimate glomerular filtration rate. *American Journal of Kidney Diseases*, 38:744–753.

[79] Limdi, J. K. and Hyde, G. M. (2003). Evaluation of abnormal liver function tests. *Postgraduate Medical Journal*, 79:307–312.

[80] Lindberg, L., Brødbæk, K., Hägerström, E. G., Bentzen, J., Kristensen, B., and Zerahn, B. (2017). Comparison of methods for estimating glomerular filtration rate in head and neck cancer patients treated with cisplatin. *Scandinavian Journal of Clinical and Laboratory Investigation*, 77:237–246.

[81] Matsuo, S., Imai, E., Horio, M., Yasuda, Y., Tomita, K., Nitta, K., Yamagata, K., Tomino, Y., Yokoyama, H., Hishida, A., Kikuchi, K., Haneda, M., Hashimoto, S., Ura, N., Ito, S., Sato, H., Watanabe, T., Ando, Y., Kusano, E., Nojima, Y., Saitoh, C., Shiigai, T., Maeda, Y., Suzuki, H., Watanabe, Y., Mitarai, T., Hasegawa, H., Fujita, T., Hosoya, T., Matsumoto, K., Fujita, T., Sanaka, T., Kanda, E., Sasaki, S., Nakano, T., Akizawa, T., Honda, H., Inoue, Y., Yoshimura, A., Fukunaga, M., Tsuji, H., Ohashi, Y., Nakajima, H., Uchida, S., Yamada, A., Umemura, S., Yanagi, M., Yamakawa, H., Kimura, K., Nishi, S., Gejyo, F., Yasuda, H., Fujikura, T., Imai, H., Kitagawa, W., Sugiyama, S., Nomura, S., Ishikawa, E., Kawabata, M., Iida, H., Wada, T., Yoshida, H., Uzu, T., Sakaguchi, M., Fukatsu, A., Nagasawa, Y., Takahara, S., Ishimura, E., Nishizawa, Y., Fukagawa, M., Umezu, M., Takeshi, N., Masaaki, I., Sugiyama, H., Makino, H., Kashihara, N., Sasaki, T., Yorioka, N., Doi, T., Kono, M., Okura, T., Tsuruya, K., Ono, A., Saito, T., Abe, Y., Furusu, A., Okuda, S., Fujimoto, S., Kunitoshi, I., Takishita, S., Ebihara, I., Shima, Y., Harada, K., Saito, Y., Numabe, A., Ishimitsu, T., Muso, E., Komiya, T., Ando, R., Hatta, T., Hotta, O., Daijo, I., Matsuyama, K., Nakano, N., Inoshita, M., Kusaka, M., and Mizuguchi, M. (2009). Revised equations for estimated GFR from serum creatinine in Japan. *American Journal of Kidney Diseases*, 53:982–992.

[82] Merion, R. M., Ashby, V. B., Wolfe, R. A., Distant, D. A., Hulbert-Shearon, T. E., Metzger, R. A., Ojo, A. O., and Port, F. K. (2005). Deceased-donor characteristics and the survival benefit of kidney transplantation. *JAMA*, 294:2726.

[83] Muggeo, V. M. (2008). segmented: an r package to fit regression models with broken-line relationships. *R News*, 8(1):20–25.

[84] Myers, G. L., Miller, W. G., Coresh, J., Fleming, J., Greenberg, N., Greene, T., Hostetter, T., Levey, A. S., Panteghini, M., Welch, M., and Eckfeldt, J. H. (2006). Recommendations

for improving serum creatinine measurement: a report from the Laboratory Working Group of the National Kidney Disease Education Program. *Clinical Chemistry*, 52:5–18.

[85] Nagai, A., Murakawa, Y., Terashima, M., Shimode, K., Umegae, N., Takeuchi, H., and Kobayashi, S. (2000). Cystatin C and cathepsin B in CSF from patients with inflammatory neurologic diseases. *Neurology*, 41:617–618.

[86] National Institute of Diabetes and Digestive and Kidney Diseases (2019). Estimating glomerular filtration rate. https://www.niddk.nih.gov/health-information/communication-programs/nkdep/laboratory-evaluation/glomerular-filtration-rate/estimating.

[87] Neuhaus, J., Jacobs, Jr, D., Baker, J., Calmy, A., Duprez, D., La Rosa, A., Kuller, L., Pett, S., Ristola, M., Ross, M., Shlipak, M., Tracy, R., and Neaton, J. (2010). Markers of inflammation, coagulation, and renal function are elevated in adults with HIV infection. *The Journal of Infectious Diseases*, 201:1788–1795.

[88] Newsome, P. N., Cramb, R., Davison, S. M., Dillon, J. F., Foulerton, M., Godfrey, E. M., Hall, R., Harrower, U., Hudson, M., Langford, A., MacKie, A., Mitchell-Thain, R., Sennett, K., Sheron, N. C., Verne, J., Walmsley, M., and Yeoman, A. (2017). Guidelines on the management of abnormal liver blood tests. *BMJ*, 67:6–19.

[89] Nyman, U., Grubb, A., Larsson, A., Hansson, L. O., Flodin, M., Nordin, G., Lindström, V., and Björk, J. (2014). The revised Lund-Malmö GFR estimating equation outperforms MDRD and CKD-EPI across GFR, age and BMI intervals in a large Swedish population. *Clinical Chemistry and Laboratory Medicine*, 52:815–824.

[90] Ocen, J., Frazer, R., Wright, S., and Bertelli, G. (2017). A comparison of measured and estimated GFR (glomerular filtration rate) for carboplatin dose calculation in women with ovarian cancer (OC). In *NCRI Cancer Conference Abstracts*. NCRI Cancer Conference.

[91] Okusa, M. D. (2010). The changing pattern of acute kidney injury: from one to multiple organ failure. In *Cardiorenal Syndromes in Critical Care*, volume 165, pages 153–158. Karger Publishers.

[92] Oliver, R. T. D., Mason, M. D., Mead, G. M., Von Der Maase, H., Rustin, G. J. S., Joffe, J. K., De Wit, R., Aass, N., Graham, J. D., Coleman, R., Kirk, S. J., and Stenning, S. P. (2005). Radiotherapy versus single-dose carboplatin in adjuvant treatment of stage I seminoma: a randomised trial. *Lancet*, 366:293–300.

[93] Ortho-Clinical Diagnostics (2008). Updated information for IDMS traceable VITROS® chemistry products CREA slides. Technical report, Ortho-Clinical Diagnostics, Rochester, NY.

[94] Ou, M., Song, Y., Li, S., Liu, G., Jia, J., Zhang, M., Zhang, H., and Yu, C. (2015). LC-MS/MS Method for serum creatinine: Comparison with enzymatic method and Jaffe method. *PLoS ONE*, 10:e0133912.

[95] Ozols, R. F., Bundy, B. N., Greer, B. E., Fowler, J. M., Clarke-Pearson, D., Burger, R. A., Mannel, R. S., DeGeest, K., Hartenbach, E. M., Baergen, R., and Mackey, D. (2003). Phase III trial of carboplatin and paclitaxel compared with cisplatin and paclitaxel in patients with

optimally resected stage III ovarian cancer: A Gynecologic Oncology Group study. *Journal of Clinical Oncology*, 21:3194–3200.

[96] Patel, S. S., Molnar, M. Z., Tayek, J. A., Ix, J. H., Noori, N., Benner, D., Heymsfield, S., Kopple, J. D., Kovesdy, C. P., and Kalantar-Zadeh, K. (2013). Serum creatinine as a marker of muscle mass in chronic kidney disease: Results of a cross-sectional study and review of literature. *Journal of Cachexia, Sarcopenia and Muscle*, 4:19–29.

[97] Peake, M. and Whiting, M. (2006). Measurement of serum creatinine–current status and future goals. *The Clinical Biochemist. Reviews*, 27:173–184.

[98] Perich, C., Ricós, C., Alvarez, V., Biosca, C., Boned, B., Cava, F., Doménech, M., Fernández-Calle, P., Fernández-Fernández, P., García-Lario, J., Minchinela, J., Simón, M., and Jansen, R. (2014). External quality assurance programs as a tool for verifying standardization of measurement procedures: Pilot collaboration in Europe. *Clinica Chimica Acta*, 432:82–89.

[99] Pianta, T. J., Endre, Z. H., Pickering, J. W., Buckley, N. A., and Peake, P. W. (2015). Kinetic estimation of GFR improves prediction of dialysis and recovery after kidney transplantation. *PLoS ONE*, 10.

[100] Poole, S. G., Dooley, M. J., and Rischin, D. (2007). Calculating carboplatin doses using the 4-variable modification of diet in renal disease (4-v MDRD) estimate of glomerular filtration rate (GFR) in the Calvert formula. *Journal of Clinical Oncology*, 25:2521.

[101] Pottel, H., Delanaye, P., Schaeffner, E., Dubourg, L., Eriksen, B. O., Melsom, T., Lamb, E. J., Rule, A. D., Turner, S. T., Glassock, R. J., De Souza, V., Selistre, L., Goffin, K., Pauwels, S., Mariat, C., Flamant, M., and Ebert, N. (2017). Estimating glomerular filtration rate for the full age spectrum from serum creatinine and cystatin C. *Nephrology Dialysis Transplantation*, 32:497–507.

[102] Pottel, H., Hoste, L., Dubourg, L., Elbert, N., Schaeffer, E., Eriksen, B. O., Melsom, T., Lamb, E. J., Rule, A. D., Turner, S. T., Glassock, R. J., De Souza, V., Selistre, L., Mariat, C., Martens, F., and Delanaye, P. (2016). An estimated glomerular filtration rate equation for the full age spectrum. *Nephrology Dialysis Transplantation*, 31:798–806.

[103] Powell, M. A., Filiaci, V. L., Rose, P. G., Mannel, R. S., Hanjani, P., DeGeest, K., Miller, B. E., Susumu, N., and Ueland, F. R. (2010). Phase II evaluation of paclitaxel and carboplatin in the treatment of carcinosarcoma of the uterus: A gynecologic oncology group study. *Journal of Clinical Oncology*, 28:2727–2731.

[104] Prado, C. M., Baracos, V. E., McCargar, L. J., Reiman, T., Mourtzakis, M., Tonkin, K., Mackey, J. R., Koski, S., Pituskin, E., and Sawyer, M. B. (2009). Sarcopenia as a determinant of chemotherapy toxicity and time to tumor progression in metastatic breast cancer patients receiving capecitabine treatment. *Clinical Cancer Research*, 15:2920–2926.

[105] Pujade-Lauraine, E., Wagner, U., Aavall-Lundqvist, E., Gebski, V., Heywood, M., Vasey, P. A., Volgger, B., Vergote, I., Pignata, S., Ferrero, A., Sehouli, J., Lortholary, A., Kristensen, G., Jackisch, C., Joly, F., Brown, C., Le Fur, N., and Du Bois, A. (2010). Pegylated liposomal doxorubicin and carboplatin compared with paclitaxel and carboplatin for patients with platinum-sensitive ovarian cancer in late relapse. *Journal of Clinical Oncology*, 28:3323–3329.

[106] Quinton, A., Lewis, P., Ali, P., Morgan, C., and Bertelli, G. (2013). A comparison of measured and estimated glomerular filtration rate for carboplatin dose calculation in stage I testicular seminoma. *Medical Oncology*, 30:1–5.

[107] R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[108] Rule, A. D., Larson, T. S., Bergstralh, E. J., Slezak, J. M., Jacobsen, S. J., and Cosio, F. G. (2004). Using serum creatinine to estimate glomerular filtration rate: accuracy in good health and in chronic kidney disease. *Annals of Internal Medicine*, 141:929–937.

[109] Rutten, I. J., van Dijk, D. P., Kruitwagen, R. F., Beets-Tan, R. G., Olde Damink, S. W., and van Gorp, T. (2016). Loss of skeletal muscle during neoadjuvant chemotherapy is related to decreased survival in ovarian cancer patients. *Journal of Cachexia, Sarcopenia and Muscle*, 7:458–466.

[110] Schwartz, G. J., Brion, L. P., and Spitzer, A. (1987). The use of plasma creatinine concentration for estimating glomerular filtration rate in infants, children, and adolescents. *Pediatric Clinics of North America*, 34:571–590.

[111] Schwartz, G. J., Haycock, G. B., Edelmann, C. M., and Spitzer, A. (1976). A simple estimate of glomerular filtration rate in children derived from body length and plasma creatinine. *Pediatrics*, 58:259–263.

[112] Schwartz, G. J., Munoz, A., Kaskel, F., Schneider, M. F., Mak, R. H., Kaskel, F., Warady, B. A., and Furth, S. L. (2009). New equations to estimate GFR in children with CKD. *Journal of the American Society of Nephrology*, 20:629–637.

[113] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464.

[114] Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components*. Wiley.

[115] Séronie-Vivien, S., Delanaye, P., Piéroni, L., Mariat, C., Froissart, M., Cristol, J.-P., and SFBC “Biology of renal function and renal failure" working group. (2008). Cystatin C: current position and future prospects. *Clinical Chemistry and Laboratory Medicine*, 46:1664–1686.

[116] Shaffi, K., Uhlig, K., Perrone, R. D., Ruthazer, R., Rule, A., Lieske, J. C., Navis, G., Poggio, E. D., Inker, L. A., and Levey, A. S. (2014). Performance of creatinine-based GFR estimating equations in solid-organ transplant recipients. *American Journal of Kidney Diseases*, 63:1007–1018.

[117] Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American Statistical Association*, 88:486–494.

[118] Shepherd, S. T., Gillen, G., Morrison, P., Forte, C., Macpherson, I. R., White, J. D., and Mark, P. B. (2014). Performance of formulae based estimates of glomerular filtration rate for carboplatin dosing in stage 1 seminoma. *European Journal of Cancer*, 50:944–952.

[119] Shi, G. P., Sukhova, G. K., Grubb, A., Ducharme, A., Rhode, L. H., Lee, R. T., Ridker, P. M., Libby, P., and Chapman, H. A. (1999). Cystatin C deficiency in human atherosclerosis and aortic aneurysms. *The Journal of Clinical Investigation*, 104:1191–1197.

[120] Shlipak, M. G., Mattes, M. D., and Peralta, C. A. (2013). Update on cystatin C: incorporation into clinical practice. *American Journal of Kidney Diseases*, 62:595–603.

[121] Telford, T., Keane, D. F., Garner, A. E., Waller, M. L., Scarsbrook, A. F., and Barnfield, M. C. (2019). Assessing the impact of inadequate hydration on isotope-gfr measurement. *Scandinavian Journal of Clinical and Laboratory Investigation*, 79:86–90.

[122] Tenstad, O., Roald, A. B., Grubb, A., and Aukland, K. (1996). Renal handling of radiolabelled human cystatin C in the rat. *Scandinavian Journal of Clinical and Laboratory Investigation*, 56:409–414.

[123] Teo, B. W., Xu, H., Wang, D., Li, J., Sinha, A. K., Shuter, B., Sethi, S., and Lee, E. J. (2011). GFR estimating equations in a multiethnic Asian population. *American Journal of Kidney Diseases*, 58:56–63.

[124] The Royal College of Radiologists (2015). Standards for intravascular contrast administration to adult patients.

[125] Uchino, S., Kellum, J. A., Bellomo, R., Doig, G. S., Morimatsu, H., Morgera, S., Schetz, M., Tan, I., Bouman, C., Macedo, E., Gibney, N., Tolwani, A., Ronco, C., and Beginning and Ending Supportive Therapy for the Kidney (BEST Kidney) Investigators (2005). Acute renal failure in critically ill patients: a multinational, multicenter study. *JAMA*, 294:813.

[126] Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition. ISBN 0-387-95457-0.

[127] Vinge, E., Lindergård, B., Nilsson–Ehle, P., and Grubb, A. (1999). Relationships among serum cystatin C, serum creatinine, lean tissue mass and glomerular filtration rate in healthy adults. *Scandinavian Journal of Clinical and Laboratory Investigation*, 59:587–592.

[128] Voskoboev, N. V., Larson, T. S., Rule, A. D., and Lieske, J. C. (2011). Importance of cystatin C assay standardization. *Clinical Chemistry*, 57:1209–1211.

[129] Webster, A. C., Nagler, E. V., Morton, R. L., and Masson, P. (2017). Chronic kidney disease. *Lancet*, 389:1238–1252.

[130] Wegiel, B., Jiborn, T., Abrahamson, M., Helczynski, L., Otterbein, L., Persson, J. L., and Bjartell, A. (2009). Cystatin C is downregulated in prostate cancer and modulates invasion of prostate cancer cells via MAPK/Erk and Androgen Receptor pathways. *PLoS ONE*, 4:e7953.

[131] Welch, M. J., Cohen, A., Hertz, H. S., Ng, K. J., Schaffer, R., Der Lijn, P. V., and White V, E. (1986). Determination of serum creatinine by isotope dilution mass spectrometry as a candidate definitive method. *Analytical Chemistry*, 58:1681–1685.

[132] Williams, E. H., Connell, C. M., Weaver, J. M. J., Beh, I., Potts, H., Whitley, C. T., Bird, N., Al-Sayed, T., Monaghan, P., Fehr, M., Cathomas, R., Bertelli, G., Quinton, A., Lewis, P., Shamash, J., Wilson, P., Dooley, M., Poole, S., Mark, P. B., Bookman, M. A., Earl, H., Jodrell, D., Tavaré, S., Lynch, A. G., and Janowitz, T. (2019). Multicentre validation of the CamGFR model for estimated glomerular filtration rate. *JNCI Cancer Spectrum*, 10.1093/jncics/pkz068.

[133] Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, 99:673–686.

[134] Wood, S. N. (2011a). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 73(1):3–36.

[135] Wood, S. N. (2011b). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73:3–36.

[136] Wood, S. N. (2017). *Generalized Additive Models: an Introduction with R*. Chapman and Hall/CRC, 2nd edition.

[137] Wright, J. G., Boddy, A. V., Highley, M., Fenwick, J., McGill, A., and Calvert, A. H. (2001). Estimation of glomerular filtration rate in cancer patients. *British Journal of Cancer*, 84:452–459.

[138] Wuerzner, G., Firsov, D., and Bonny, O. (2014). Circadian glomerular function: From physiology to molecular and therapeutical aspects. *Nephrology Dialysis Transplantation*, 29:1475–1480.

[139] Zhang, P. (1993). Model selection via multifold cross validation. *The Annals of Statistics*, 21:299–313.

[140] Zweifel, M., Jayson, G. C., Reed, N. S., Osborne, R., Hassan, B., Ledermann, J., Shreeves, G., Poupard, L., Lu, S. P., Balkissoon, J., Chaplin, D. J., and Rustin, G. J. S. (2011). Phase II trial of combretastatin A4 phosphate, carboplatin, and paclitaxel in patients with platinum-resistant ovarian cancer. *Annals of Oncology*, 22:2036–2041.

Appendix A

I have been involved in several other projects for which I have contributed statistical analysis. While these are not discussed in detail in the body of the thesis, they are included in this appendix for completeness. I was the statistician on an analysis of trends in T-cell checkpoint-targeted cancer immunotherapy clinical trials, and on the analysis of interim results from the national Hepatitis Prevention, Control, and Elimination (HPCE) in Mongolia. This work contributed to the publication of two additional papers and a further paper which is currently under review. These papers are presented below.

Regarding these facts, the recently issued guidelines from the RECIST working group is of major importance to standardize the i-RECIST. These guidelines (i) define the concept of "unconfirmed PD '/' confirmed PD" according to the results of the imaging performed after the initial PD, (ii) include clinical stability in the definition, and (iii) define confirmed PD when a new lesion occurs [5]. The authors should be applauded for this initiative; the definition of endpoints do not differ according to the sponsor.

Use of a common and standardized language is critical when designing and analyzing clinical trials. Current ongoing trials used widely different definitions of confirmed PD, creating major difficulties when discussing results. The RECIST working group guidelines should be endorsed to avoid non-reproducibility of trial results.

J. Le Lay[1], H. Jarraya[2], L. Lebellec[3] & N. Penel[1,4*]

Departments of [1]Clinical Research and Innovation; [2]Imaging Centre Oscar Lambret, Lille; [3]Department of Cancerology, Medical School, Lille University, Lille; [4]Medical Oncology Department, Centre Oscar Lambret, Lille, France

(*E-mail: n-penel@o-lambret.fr)

Disclosure

The authors have declared no conflicts of interest.

References

1. Eisenhauer EA, Therasse P, Bogaerts J et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1·1). Eur J Cancer 2009; 45: 228–247.
2. Benjamin RS, Choi H, Macapinlac HA et al. We should deists using RECIST, at least in GIST. J Clin Oncol 20017; 25: 1760–1764.
3. Chiou VL, Burotto M. Pseudoprogression and immune-related response in solid tumors. J Clin Oncol 2015; 33: 3541–3543.
4. Borghaei H, Paz-Ares L, Horn L et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. N Engl J Med 2015; 373: 1627–1639.
5. Seymour L, Bogaerts J, Perrone A et al. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. Lancet Oncol 2017; 18: e143–e152.

doi:10.1093/annonc/mdx168
Published online 5 April 2017

Funding

None declared.

Cancer immunotherapy trial registrations increase exponentially but chronic immunosuppressive glucocorticoid therapy may compromise outcomes

T cell checkpoint-targeted immunotherapy is effective in multiple cancers, but only in subsets of patients [1]. Failure of immunotherapy may be secondary to tumour intrinsic and/or systemic factors that impair immune response. Glucocorticoid administration has known systemic immunosuppressive effects [2] with potential to impair immunotherapy outcome [3], and should therefore be regulated at patient enrolment.

We performed a cross-sectional analysis of T cell checkpoint-targeted cancer immunotherapy trials in solid malignancies registered on the U.S. National Institutes of Health (NIH) trial registry (clinicaltrials.gov) by October 7, 2016. Trials were searched by study type, condition, and interventions targeting the T cell checkpoint proteins CTLA-4, PD-1, PD-L1, PD-L2, LAG3, B7-H3, CD137, OX40, CD27 and GITR. Trials were reviewed manually and independently by two clinicians and registered data on glucocorticoid administration within enrolment criteria recorded.

We identified 1017 registered T cell checkpoint-targeted cancer immunotherapy trials. The number of registrations has progressively increased, exponentially between 2010 and 2015 ($R^2 = 0.95$) (Figure 1A). For the completed years, 2001–2015, chronic glucocorticoid administration was stated as an exclusion criterion in 40% (276/685), permitted in 29% (201/685) and not specified in 30% (208/685) of trial registration details. The proportion of trials that did not allow glucocorticoid use has decreased significantly ($P < 0.001$), while the proportion allowing glucocorticoid use has increased significantly ($P < 0.001$) (Figure 1B). Of the trials permitting glucocorticoid use, the maximum permitted dose of prednisolone equivalent per day was up to 10 mg in 57% of trials (115/201), over 10 mg in 4% of trials (9/201) and not specified in 14% of trials (28/201); 24% of trials (49/201) permitted chronic glucocorticoid use for physiological replacement.

These findings are concerning. The immunosuppressive effects of glucocorticoids are dose dependent, starting at less than 10 mg of prednisolone per day [2], and may be compounded by hypoalbuminaemia present in patients with cancer [4]. Moreover, our pre-clinical work has demonstrated that low dose glucocorticoid administration is sufficient to suppress response to cancer immunotherapy [3]. Therefore, unregulated glucocorticoid administration may result in treatment failure independent of the T cell checkpoint-targeted agent or tumour type.

The use of glucocorticoids as appetite stimulants and antiemetics, particularly relevant in combination trials with emetogenic chemotherapy or radiotherapy, may also be immunosuppressive and will require critical review. While the use of glucocorticoids for adrenal replacement or chronic immune illness may be unavoidable, stratification for their use at enrolment should be considered.

In addition to glucocorticoid administration, endogenous glucocorticoid levels may also impact on response to immunotherapy. Monitoring and stratification according to baseline glucocorticoid levels, or clinical surrogates of these such as longitudinal weight measurements [3], may yield predictive and prognostic markers of response.

We note that the chronic use of glucocorticoids should be considered independently from the use of glucocorticoids in managing immune-related adverse events during immunotherapy. In

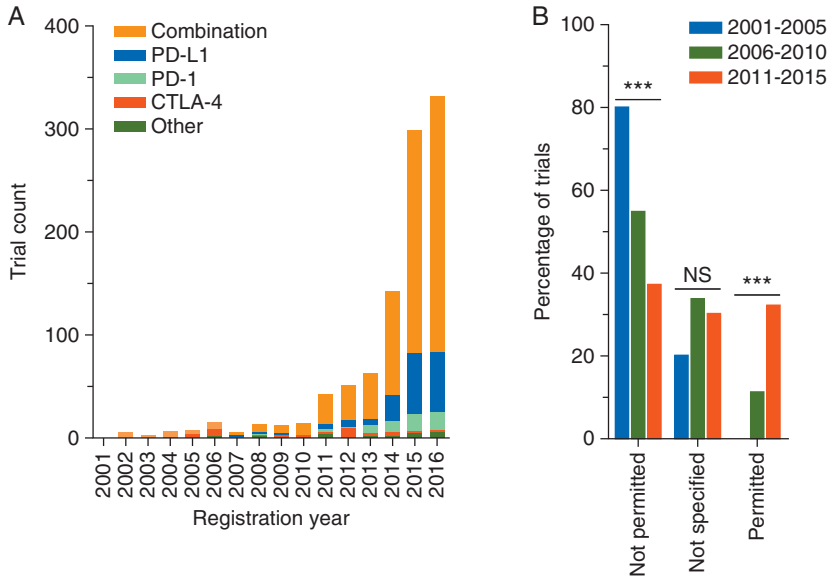


Figure 1. Longitudinal registration count and glucocorticoid administration in T cell checkpoint-targeted cancer immunotherapy trials. (A) Annual registration count of T cell checkpoint-targeted cancer immunotherapy trials. Trials registered by October 7, 2016 on the U.S. National Institutes of Health (NIH) trial registry were categorized according to year of registration and checkpoint protein target. T cell checkpoint proteins targeted in fewer than 10 trials are grouped as 'other', and include single agent OX40, GITR, CD137, B7-H3, LAG3, PD-L2, CD27 and trials comparing multiple checkpoint-targeting agents. 'Combination' trials include all pre-defined T cell checkpoint-targeted agents used in combination with another agent. (B) Specification of glucocorticoid administration within enrolment criteria of T cell checkpoint-targeted cancer immunotherapy trials. Trials registered on the U.S. NIH trial registry between 2001 and 2015 were categorized according to the specification of chronic systemic glucocorticoid administration within registered patient enrolment criteria. Univariate analysis for data presented was performed using the Cochran–Armitage test for trend assuming monotonical change over time and expected frequencies were met (80% of expected frequencies >5). ***$P < 0.001$; NS= not significant.

fact, the positive correlation of autoimmune side effects and treatment efficacy [5] provides further rationale for considering the role of systemic immunomodulatory variables in determining response to immunotherapy.

Our study is limited by the exclusive use of data from the U.S. NIH trial registry. However, this is the largest clinical trial registry, and the registration of key inclusion and exclusion criteria is international standard [6] and has been mandatory for consideration of publication by the International Committee of Medical Journal Editors member journals since 2005.

In summary, we find glucocorticoid administration to be a neglected immunomodulatory variable in cancer immunotherapy trials, and suggest striving for greater harmony in the monitoring and regulation of systemic glucocorticoids to improve outcomes in cancer immunotherapy.

C. M. Connell[1,2†], S. Raby[1†], I. Beh[1], T. R. Flint[2], E. H. Williams[2], D. T. Fearon[2,3,4], D. I. Jodrell[1,2] & T. Janowitz[1,2*]

[1]Department of Oncology, University of Cambridge, Addenbrooke's Hospital, Cambridge; [2]Cancer Research UK Cambridge Institute, University of Cambridge Li Ka Shing Centre, Cambridge, UK; [3]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York; [4]Weill Cornell Medical College, New York, USA

(*E-mail: tj212@cam.ac.uk)

[†]Both authors contributed equally

Funding

This work was supported by the Wellcome Trust Translational Medicine and Therapeutics Programme [RJAG/076 to TJ], Cancer Research UK and the Cambridge Translational Medicine and Therapeutics Academic Clinical Fellowship Programme (to CMC).

Disclosure

The authors have declared no conflicts of interest.

References

1. Topalian SL, Hodi FS, Brahmer JR et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. N Engl J Med 2012; 366(26): 2443–2454.
2. Wolfe F, Caplan L, Michaud K. Treatment for rheumatoid arthritis and the risk of hospitalization for pneumonia: associations with prednisone, disease-modifying antirheumatic drugs, and anti-tumor necrosis factor therapy. Arthritis Rheum 2006; 54(2): 628–634.
3. Flint TR, Janowitz T, Connell CM et al. Tumor-induced IL-6 reprograms host metabolism to suppress anti-tumor immunity. Cell Metab 2016; 24(5): 672–684.
4. Lewis G, Jusko W, Burke C, Graves L. Prednisone side-effects and serum-protein levels. Lancet 1971; 298(7728): 778–781.
5. Downey SG, Klapper JA, Smith FO et al. Prognostic factors related to clinical response in patients with metastatic melanoma treated by CTL-associated antigen-4 blockade. Clin Cancer Res 2007; 13(22 Pt 1): 6681–6688.
6. Zarin DA, Tse T, Williams RJ, Carr S. Trial reporting in clinicaltrials.gov—the final rule. N Engl J Med 2016;375: 1998–2004.

doi:10.1093/annonc/mdx181
Published online 14 April 2017

The Oncologist®

Brief Communications

Cancer Immunotherapy Trials Underutilize Immune Response Monitoring

Claire M. Connell,[a,b] Sophie E.M. Raby,[a] Ian Beh,[a] Thomas R. Flint,[b] Edward H. Williams,[b] Douglas T. Fearon,[b,c,d] Duncan I. Jodrell,[a,b] Tobias Janowitz[a,b]

[a]Department of Oncology, University of Cambridge, Addenbrooke's Hospital, Cambridge, United Kingdom; [b]Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Cambridge, United Kingdom; [c]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA; [d]Weill Cornell Medical College, New York, USA

Disclosures of potential conflicts of interest may be found at the end of this article.

Abstract

Immune-related radiological and biomarker monitoring in cancer immunotherapy trials permits interrogation of efficacy and reasons for therapeutic failure. We report the results from a cross-sectional analysis of response monitoring in 685 T-cell checkpoint-targeted cancer immunotherapy trials in solid malignancies, as registered on the U.S. National Institutes of Health trial registry by October 2016. Immune-related radiological response criteria were registered for only 25% of clinical trials. Only 38% of trials registered an exploratory immunological biomarker, and registration of immunological biomarkers has decreased over the last 15 years. We suggest that increasing the utilization of immune-related response monitoring across cancer immunotherapy trials will improve analysis of outcomes and facilitate translational efforts to extend the benefit of immunotherapy to a greater proportion of patients with cancer. ***The Oncologist*** 2018;23:116–117

T-cell checkpoint-targeted cancer immunotherapies are making an increasing impact on clinical practice, and their investigation in clinical trials has risen exponentially [1]. Effective and efficient response evaluation is essential, but presents challenges due to idiosyncratic radiological responses and lack of early response biomarkers. Concerted efforts to overcome these challenges have included the generation of immune-related response criteria (irRC) [2, 3] to better accommodate immunotherapy-associated response kinetics, and recommendations for immunological monitoring throughout all phases of clinical trial design to help identify predictive/prognostic biomarkers of response and mechanistic insight into patterns of resistance [3, 4].

We assessed quantitatively the inclusion of irRC and immunological biomarker primary response monitoring in the registration details of T-cell checkpoint-targeted cancer immunotherapy trials in solid malignancies registered on the NIH trial registry (clinicaltrials.gov) by October 7, 2016. To overcome potential limitations from incomplete registrations, data were analyzed across three consecutive 5-year periods.

During the completed years, 2001–2015, 91% (622/685) of immunotherapy trials registered use of a radiological endpoint. Analyzing the trend of utilization, we found a significantly increasing ($p = .014$) proportion of these trials specified use of the World Health Organization/Response Evaluation Criteria in Solid Tumors, reaching 56% (307/551) in the last 5 years. Use of the irRC (established in 2009) also increased, but only reached 25% (135/551) of trials. A total of 38% (234/622) of

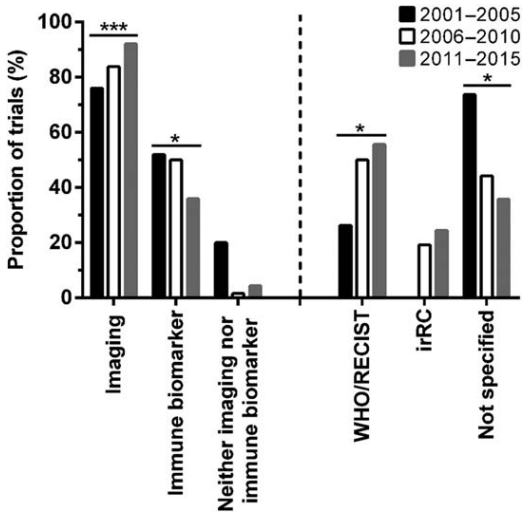


Figure 1. Radiological and immune biomarker monitoring in T-cell checkpoint-targeted cancer immunotherapy trials. All 685 T-cell checkpoint-targeted cancer immunotherapy trials registered on the U.S. NIH trial registry between 2001 and 2015 were categorized according to year of registration and registered imaging and immune biomarker outcome measures (left) and radiological response criteria (right). Six trials used the Response Assessment in Neuro-Oncology criteria (not shown). Immune biomarkers included monitoring of any immune-related parameter, including via immunohistochemistry, immune cell counts, cytokine analysis, and humoral/cellular immune responses. Univariate analysis for data presented was performed using the Cochran-Armitage test for trend. *, $p < .05$; ***, $p < .001$.

Abbreviations: irRC, immune-related response criteria; RECIST, Response Evaluation Criteria in Solid Tumors; WHO, World Health Organization.

Correspondence: Tobias Janowitz, M.B., B.Chir., Ph.D., Telephone: +44 1223 769786; e-mail: tj212@cam.ac.uk Received May 15, 2017; accepted for publication August 22, 2017; published Online First on October 11, 2017. http://dx.doi.org/10.1634/theoncologist.2017-0226

trials registered use of an imaging endpoint without specifying imaging criteria. Overall, only 38% (259/685) of trials included immunological biomarker monitoring, and this proportion decreased significantly ($p = .011$) over three consecutive 5-year periods (Figure 1).

These findings highlight three areas of concern. Firstly, underutilization of the irRC may be associated with an underestimation of treatment response [2, 5]. This impact may be dependent on tumor type or checkpoint-targeted agent [6]. However, prospective inclusion of radiological criteria that accommodate unconventional response kinetics is essential in order to avoid future misclassification of response and inappropriate cessation of effective therapy. Secondly, the failure to incorporate immunological, and emerging host-centric [7], biomarker monitoring presents a missed opportunity to establish surrogate markers of response and/or resistance [3, 4]. Thirdly, specification and/or metrics of imaging outcome measures were found wanting in 234 trials, an omission that should be addressed, particularly in view of recent regulations [8].

The underutilization of recommended radiographic and immunological monitoring identified by these data, together with suboptimal dose finding identified by others [9], may limit the interpretation of clinical trial results and thereby the development of effective cancer immunotherapies and consensus guidelines for their use. Increasing the incorporation of these measures will help answer the call for greater accountability in the design of clinical trials to optimize the value of data generated [4, 8]. This will improve our knowledge, and application, of immunotherapy for the benefit of patients with cancer.

Acknowledgments

This work was supported by the Wellcome Trust Translational Medicine and Therapeutics Programme [RJAG/076 to T.J.], by Cancer Research UK [C42738/A24868 to T.J.], and the Cambridge Translational Medicine and Therapeutics Academic Clinical Fellowship Programme (C.M.C.).

Disclosures

The authors indicated no financial relationships.

References

1. Connell CM, Raby S, Beh I et al. Cancer immunotherapy trial registrations increase exponentially but chronic immunosuppressive glucocorticoid therapy may compromise outcomes. Ann Oncol 2017;28: 1678–1679.

2. Wolchok JD, Hoos A, O'Day S et al. Guidelines for the evaluation of immune therapy activity in solid tumors: Immune-related response criteria. Clin Cancer Res 2009;15:7412–7420.

3. Hoos A, Eggermont AM, Janetzki S et al. Improved endpoints for cancer immunotherapy trials. J Natl Cancer Inst 2010;102:1388–1397.

4. Mellman I, Hubbard-Lucey VM, Tontonoz MJ et al. De-risking immunotherapy: Report of a consensus workshop of the Cancer Immunotherapy Consortium of the Cancer Research Institute. Cancer Immunol Res 2016;4:279–288.

5. Hodi FS, Hwu WJ, Kefford R et al. Evaluation of immune-related response criteria and RECIST v1.1 in patients with advanced melanoma treated with pembrolizumab. J Clin Oncol 2016; 34:1510–1517.

6. Nishino M, Ramaiya NH, Chambers ES et al. Immune-related response assessment during PD-1 inhibitor therapy in advanced non-small-cell lung cancer patients. J Immunother Cancer 2016;4:84.

7. Flint TR, Fearon DT, Janowitz T. Connecting the metabolic and immune responses to cancer. Trends Mol Med 2017;23:451–464.

8. Zarin DA, Tse T, Williams RJ et al. Trial reporting in clinicaltrials.gov — The final rule. N Engl J Med 2016;375:1998–2004.

9. Tosi D, Laghzali Y, Vinches M et al. Clinical development strategies and outcomes in first-in-human trials of monoclonal antibodies. J Clin Oncol 2015; 33:2158–2165.

Interim Results of the Hepatitis C Elimination Program in Ulaanbaatar, Mongolia

Naranbaatar Dashdorj[1,10,12], Harry Potts[2,10], Andreas Bungert[1,10], Purevjargal Bat-Ulzii[1,3], Bekhbold Dashtseren[1,3,4], Uurtsaikh Baatarsuren[1], Maral Myanganbayar[1], Edward H. Williams[5], Jeffrey S. Glenn[6], Zulkhuu Genden[1,3], Altankhuu Mordorj[1,3], Davaadorj Duger[4], Tsolmon Unurjargal[4,7], Dahgwahdorj Yagaanbuyant[1,3,4], Tobias Janowitz[8,9,11,12], Naranjargal Dashdorj[1,3,11,12]

[1] Onom Foundation, Ulaanbaatar, Mongolia
[2] School of Clinical Medicine, University of Cambridge, Cambridge, United Kingdom
[3] Liver Center, Ulaanbaatar, Mongolia
[4] Mongolian National University of Medical Sciences, Ulaanbaatar, Mongolia
[5] Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, United Kingdom
[6] Stanford University School of Medicine, Stanford, CA, USA
[7] University Hospital, Ulaanbaatar, Mongolia
[8] Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, NY 11724, United States of America
[9] Northwell Health, New York, 11021, United States of America
[10] These authors contributed equally as first authors.
[11] These authors contributed equally as last authors.
[12] Corresponding authors:
nd@onomfoundation.org; janowitz@cshl.edu; dashdorj@onomfoundation.org

Abstract

Background

Mongolia, the country with the highest prevalence of viral hepatitis and the highest liver cancer mortality in the world, initiated a national Hepatitis Prevention, Control, and Elimination (HPCE) Program in 2016. This program aims to eliminate hepatitis C virus (HCV) nationwide by the end of 2020, ahead of the World Health Organization global target of 2030. This study quantifies the impact of the HPCE Program on the prevalence of HCV infection in Mongolia's capital Ulaanbaatar, home to 46% of the country's population.

Methods

We performed two randomized cross-sectional surveys in the adult population of Ulaanbaatar. The first survey was conducted in 2013 and the second in 2018, prior to and after implementation of the HPCE Program respectively. Anti-HCV antibody (HCV antibody) positivity as a marker of HCV exposure and HCV-RNA detectability as a marker of disease activity were determined.

Results

Screening for HCV was conducted in 571 individuals in 2013 and in 579 individuals in 2018. HCV antibody positivity remained unchanged between 2013 and 2018, at 7.7% and 7.8% respectively. The fraction of HCV antibody positive individuals with detectable HCV-RNA decreased from 79.1% in 2013 to 35.3% in 2018. The largest absolute decrease was seen in women over the age of 50 years.

Conclusions

Between 2013 and 2018, the burden of active HCV infection in the adult population of Ulaanbaatar has fallen markedly, in concordance with the intent of the HPCE program. HCV elimination in Mongolia appears to be a realistic goal.